This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2026/10/24/the-certificate-that-says-nothing/.

The Certificate That Says Nothing

Heather Grizzle · October 24, 2026

A sign language AI vendor's ISO 42001 certificate can be genuine and still say nothing about whether Deaf people understand the output.

Inhalt

1. Signed language AI is about to become assurable. What a certificate will not tell you is whether Deaf people can understand the output.
2. A note on where we are writing from
3. Two different things called assurance
4. What became real in 2025 and 2026
5. What is not yet real
6. The test that actually discriminates
7. Who is entitled to give the assurance
8. When the output is wrong
9. Five questions, and our own answers
10. The certificate is real, and so is the gap

Signed language AI is about to become assurable. What a certificate will not tell you is whether Deaf people can understand the output.

NCG Insights October 2026

Consider a scene likely to become routine within the next few years. A procurement officer in a public body, university, healthcare system or large regulated firm asks a supplier of signed language AI a perfectly reasonable question, which is whether the supplier can demonstrate that the system is safe, accurate and responsibly built. The

supplier produces a certificate. It may be entirely legitimate, issued by a competent certification body against a published international standard and supported by an accreditation infrastructure now being built specifically for artificial intelligence management systems. The procurement officer reads it as evidence that somebody independent has examined the supplier's AI practices, places it in the file and moves on. The certificate may be perfectly genuine and still provide no evidence whatsoever that a Deaf person watching the system's signed output can understand it.

That distinction matters because AI assurance is moving rapidly from an emerging professional practice into institutional infrastructure. The international standards needed to certify AI management systems now exist, as do the standards governing accreditation of the bodies that conduct those audits, and governments are investing in assurance tools, professional competencies and a market of third-party providers. What we have not been able to identify anywhere in that apparatus is an equivalent, publicly available, independently validated instrument capable of establishing that signed language AI output is linguistically comprehensible, contextually appropriate and fit for the Deaf people expected to rely on it.

That is the gap examined here. It is not an argument against AI assurance and not a claim that management-system certification is meaningless, since the emerging infrastructure does something genuinely useful and the problem it creates does not require anybody to act in bad faith. The failure mode here is not fraud but inference.

A note on where we are writing from

Novara Consulting Group is writing as an organisation that expects its own work to be scrutinised. We do not intend to enter the certification market, and this article should not be read as a prospectus for doing so. Our interest is in the evidence a certificate cannot supply, and any work we do on that evidence is subject to the same questions we put to everyone else near the end of this piece. We are publishing this article because we would rather that difficult assurance requirements exist before somebody applies them to us than construct requirements afterwards that conveniently describe what we already do well.

The questions near the end are therefore questions we have put to ourselves, with our own answers published alongside them. Several of those answers are inadequate, and that is deliberate. An assurance proposition its author can only publish after passing it is not functioning as an assurance proposition. It is functioning as marketing.

Two different things called assurance

The word assurance is beginning to carry several meanings in the AI market, and signed language technology makes the differences unusually consequential. The first is process assurance, which asks whether an organisation has a functioning management system around its AI activities: whether risks are identified, responsibilities assigned, decisions documented, controls implemented and incidents handled, and whether any mechanism exists through which evidence can cause a decision, control or deployment to change. This is familiar territory in conformity assessment, because management-system certification has long served the same function in quality, environmental management and information security.

The second is outcome assurance, which asks whether the system performs adequately for the people expected to rely on it. In signed language AI that is not an abstract question about whether a model produces something recognisably sign-like. It is whether Deaf viewers understand the output, whether meaning survives the transformation, whether grammatical information carried through space, movement, facial expression and other non-manual features is preserved, whether register suits the context, whether consequential information is omitted or distorted, and whether performance holds across the population and the situations in which the technology will actually be deployed.

A third question cannot be collapsed into either category, which is whether the linguistic material the technology was built from was legitimately obtained and remains legitimately used. Signed language corpora and segment libraries are not abstract collections of tokens. They contain recordings, representations and derived data originating from identifiable human beings and from language communities whose linguistic and cultural interests cannot be reduced to an organisation′s ownership of a video file. Provenance, consent, licensing, secondary use, withdrawal, representation and benefit-sharing therefore belong inside the assurance conversation, even where a conventional audit can establish only that procedures governing them exist.

The conformity-assessment distinction underneath all of this is the critical one. Certification against ISO/IEC 42001 concerns an organisation′s artificial intelligence management system, and ISO/IEC 42006 strengthens the infrastructure around the bodies auditing and certifying those management systems. Neither changes the object being certified. A supplier could maintain a genuinely mature AI management system, undergo a competent independent audit and obtain wholly legitimate certification

while the certificate still tells a buyer nothing about whether a particular Deaf viewer can understand a particular output. There is no contradiction in any of that. The error occurs only when evidence of the first is treated as evidence of the second, and that is where inference does its damage. A buyer sees evidence of process assurance, assumes it carries an outcome claim the certificate was never designed to make, and stops asking. Every document in the chain can be authentic and every organisation involved can have behaved impeccably, and the Deaf person at the end of it can still receive signed output that no qualified Deaf assessor ever established was comprehensible.

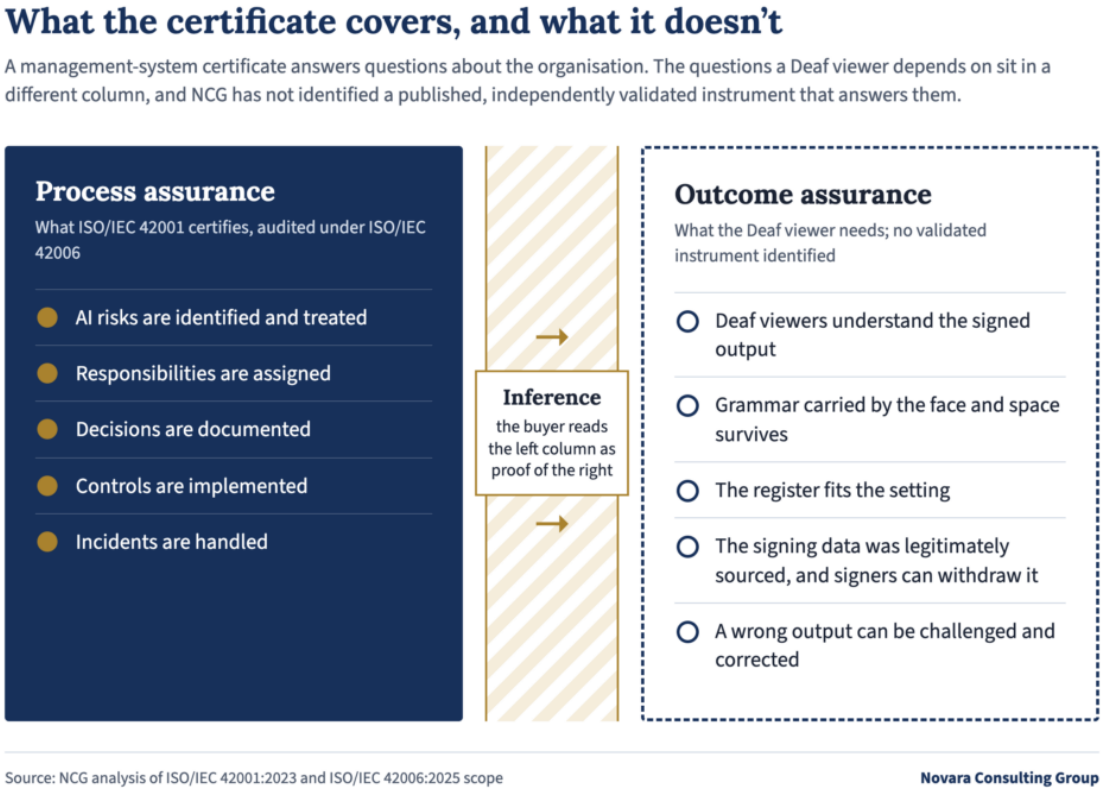

Figure 1. The certificate and the gap. A management-system certificate evidences process assurance; it does not evidence what a Deaf viewer understands.
Novara Consulting Group

What became real in 2025 and 2026

It is worth being precise about the machinery, because imprecision is exactly what allows different kinds of assurance to become interchangeable in a procurement conversation. ISO/IEC 42001 establishes requirements for an artificial intelligence manage-

ment system, giving organisations that develop, provide or use AI a structure through which AI-related responsibilities, risks, controls and continual improvement can be governed. Its importance is substantial and its scope needs to stay visible whenever a certificate is presented as evidence, because it is management-system certification and it does not become a product-specific linguistic performance evaluation merely because a buyer would find that more useful.

A significant further piece of the infrastructure arrived on 7 July 2025 with the publication of ISO/IEC 42006:2025, which specifies additional requirements for bodies auditing and certifying AI management systems against ISO/IEC 42001, supplementing ISO/IEC 17021-1, and supports the demonstration of competence, consistency and reliability by those certification bodies. ISO indicates that it can also be integrated into broader conformity-assessment schemes for AI products, processes or services under ISO/IEC 17065, so it should not be characterised as confined to management-system certification. The narrower and more important point holds under either scheme, which is that ISO/IEC 42006 governs the competence of certification bodies and supplies no methodology for evaluating signed language linguistic performance. The distinction is easy to lose because accreditation increases confidence in certification and is designed to do exactly that. What accreditation does not do is silently expand the scope of the underlying certificate. A more trustworthy audit of an AI management system remains an audit of an AI management system, and it does not become a linguistic validation study because the auditor is better qualified.

The United Kingdom has meanwhile been deliberately constructing a broader assurance ecosystem. The Trusted third-party AI assurance roadmap, published on 3 September 2025, committed the government to convening a multistakeholder consortium working towards a future AI assurance profession, developing a skills and competencies framework for AI assurance, and establishing an AI Assurance Innovation Fund to develop novel assurance solutions across the industrial strategy sectors. That consortium launched the following year, chaired by BCS with techUK leading its industry advisory group, and its first-year mandate covers a voluntary code of ethics for practitioners, a competency framework and a map of the information access assurance providers require from their clients.

The commercial ambition is substantial. The roadmap describes a market of over 524 companies contributing approximately £1.01 billion in gross value added in 2024, with the potential to exceed £18.8 billion by 2035 if barriers to widespread AI adopti-

on are addressed, and the ecosystem is now reported to employ over 12,000 people. AI assurance is being built as a serious professional market rather than an informal ethics exercise, and nothing in this article depends on pretending otherwise.

The roadmap is also candid about the quality problem within the market it seeks to develop. It reports uncertainty about the quality of existing AI assurance certifications, observing that at the time of writing none were issued by UKAS-accredited organisations, and notes that the quality infrastructure surrounding assurance providers remains under development. It cites research by the World Privacy Forum finding that as many as 38 per cent of the AI governance tools assessed mention, recommend or incorporate problematic metrics capable of producing harmful results. That accreditation position has since moved, and it moved in the direction the roadmap intended: UKAS has progressed from piloting accreditation for ISO/IEC 42001 certification to granting it, and accreditation bodies including ANAB in the United States have done the same, so accredited certification against the management-system standard is now genuinely available. This strengthens rather than weakens the argument here. The accredited infrastructure that did not exist when the roadmap was written now exists, and it still certifies management systems.

International AI testing infrastructure continues to develop alongside it. ISO/IEC TS 42119-2:2025, covering the testing of AI systems, was published in November 2025 as a Technical Specification, while the part of the series addressing verification and validation analysis had reached final text without publication as of early 2026, and further parts on red teaming and generative systems remain in earlier stages. That work makes the absence of a corresponding signed language measurement layer more consequential rather than less. The question this article asks is what happens as a rapidly maturing assurance infrastructure reaches a linguistic domain in which the corresponding domain-specific measurement layer remains substantially underdeveloped.

What is not yet real

Begin with auditor competence. General AI auditor competency can establish whether an auditor understands AI risk, management systems, controls, evidence, governance and the requirements of the applicable conformity-assessment regime. Those are real competencies and none of them confers the ability to determine whether signed language output is linguistically correct. An auditor who does not know the signed language being evaluated cannot independently assess the output, and what they can do

instead is inspect whether the supplier has a procedure for conducting that assessment, whether it was followed and whether records exist. Those are useful findings about a process rather than findings about the language.

That creates a second problem, which is what the underlying linguistic assessment should measure. Signed languages encode meaning through mechanisms that do not map neatly onto written-language translation metrics, since spatial relationships, movement, orientation, facial expression, mouth actions and other non-manual features can all carry grammatical or semantic information. Gloss-based representations are useful research artefacts and they do not capture the complete linguistic event, so a score derived from an intermediate representation can tell us something about system behaviour while telling us nothing about whether Deaf people understood the final output correctly.

Research on signed language generation, recognition, translation and evaluation is active, and the claim here is not that nobody has ever evaluated signed language technology, because researchers, Deaf participants, linguists and developers plainly have. The missing element is standardisation sufficient for assurance. A buyer needs to know what was measured, who performed the assessment, how raters were selected, what the denominator was, which error categories were used, how disagreement was handled, whether results were stratified by signer and by relevant linguistic variables, which deployment contexts were tested, and whether the instrument itself has ever been validated. A supplier reporting a single accuracy figure without those answers has provided a number rather than an assurance claim.

The community-governance terrain presents a further problem. The World Federation of the Deaf and the World Association of Sign Language Interpreters have published a position on signing avatars cautioning against substitution for human signers and raising concerns about linguistic quality and appropriate use. Technology has advanced since that position was first issued and the underlying governance question has not dissolved with it. A technical assurance regime cannot determine that the political, linguistic and cultural concerns of an affected community have evaporated because the engineering improved. Those concerns have to enter the decision architecture somewhere, and somebody has to hold enough authority for an objection to matter.

Demand makes the distinction commercially significant. Legal and policy recognition of signed language access is increasing across several jurisdictions, including through

the British Sign Language Act 2022 and the planning structure established under the British Sign Language (Scotland) Act 2015. At the same time the technical accessibility standards most familiar to procurement teams do not create a straightforward mandatory floor for signed language provision in every digital context, since signed language interpretation for prerecorded audio content sits at WCAG success criterion 1.2.6, Level AAA, rather than Level AA. That combination produces precisely the environment in which certification acquires significance beyond its formal scope. Institutions face pressure to improve signed language access, emerging technology appears able to supply it at scale, technical requirements do not answer every deployment question, and procurement teams need defensible evidence of due diligence. A recognised certificate is extraordinarily attractive under those conditions. The danger is not the certificate but the act of asking it to answer a question it was never constructed to answer.

The test that actually discriminates

There is a question that cuts through a remarkable amount of governance language, which is what the governance has actually stopped. Not what the framework covers, how many committees exist, how many controls appear in a spreadsheet, or whether there is an AI policy, ethics statement or advisory group, but the point at which governance encountered a proposed action and caused something materially different to happen. A dataset rejected because its provenance could not be established. A release delayed because Deaf evaluators identified an unacceptable error pattern. A deployment context excluded because the consequences of mistranslation were too high. Consent terms rewritten, an evaluation failed, a performance claim removed because the evidence could not support it. Records of that kind are valuable because they demonstrate that governance has causal force.

This does not require a document literally titled Non-Conformity Log, since mature organisations may record adverse decisions through corrective-action systems, risk-treatment records, internal audit findings, issue registers, release gates or incident systems. The label matters far less than the evidence trail, and what matters is whether the organisation can show that its governance has ever produced a decision that was inconvenient. An assurance pack containing policies, committee structures, risk registers and control descriptions tells a buyer what is supposed to happen. Evidence

of rejected data, failed tests, blocked releases and overturned decisions tells the buyer what happened when the system came under pressure.

The principle has a direct parallel in our research on Deaf education data, where the absence of evidence is frequently the finding itself. Where an institution cannot produce the record needed to determine who is being served, how funding reaches them, what outcomes are measured or who is accountable, the responsible conclusion is not that something is wrong but that the evidence necessary to determine what is happening is unavailable. AI governance deserves identical treatment. If an organisation says governance constrains its AI programme, ask for evidence of the constraint operating.

Who is entitled to give the assurance

The hardest problem in signed language AI assurance may prove not to be technical at all. Signed language datasets are made from people. Video corpora contain identifiable signers, segment libraries preserve individual performances that can later be recombined into outputs the original signer never produced, and derived representations can remain connected to source material in ways that are not obvious to the person who agreed to be recorded. Commercial arrangements change, models change, intended uses change, organisations are acquired, and a corpus gathered for one purpose becomes valuable for another.

An assurance regime therefore has to ask considerably more than whether a consent form exists. It has to ask what the person agreed to, whether downstream AI use was contemplated, whether commercial reuse was disclosed, whether withdrawal is possible and what withdrawal actually removes, whether derivatives can be traced, who controls subsequent licensing, whether the represented signing population matches the population for whom performance claims are made, and whether communities whose linguistic resources generate commercial value hold any meaningful authority over that use.

A conventional third-party auditor can inspect much of this, examining records, determining whether procedures exist, sampling evidence and establishing whether an organisation followed its own rules. What an auditor cannot manufacture through professional independence is community legitimacy, which is why the distinction between participation and authority carries so much weight. Consultation means Deaf people were asked. Participation may mean Deaf people worked on the project. Em-

ployment may mean Deaf people were paid to contribute linguistic expertise, translation or data. Advisory structures may mean Deaf people had a formal route for raising concerns. None of those facts, standing alone, establishes decision authority, which answers a different question entirely: who can say no, and what happens when they do. If a Deaf advisory body objects to a deployment and the supplier proceeds unchanged, the body is advisory. If a Deaf governance body can withhold approval and the release cannot proceed until the objection is resolved, the body holds authority. Both arrangements may have value and they should not be described as equivalent.

Figure 2. Consultation is not authority. Only a body that can withhold approval changes what ships. Novara Consulting Group

This is where second-party assurance deserves more attention than it currently receives. Third-party assurance is powerful where the question is conformity against an established standard. Signed language technology additionally raises questions for which the affected linguistic community is not merely another source of evidence but a party with a legitimate interest in determining acceptable use. Deaf-led bodies there-

fore need roles extending beyond being interviewed by auditors or recruited as test participants, and in at least some parts of the assurance architecture they need actual decision rights. That arrangement is uncommon for an obvious reason. Real authority can stop things, which is precisely why it counts.

When the output is wrong

A further question becomes visible only after a system leaves the laboratory, the procurement process and the assurance audit and reaches the Deaf person expected to use it, which is what happens when that person says the output was wrong. A system could have documented governance, legitimate data provenance, Deaf evaluators, competent third-party auditors and a formal approval structure and still fail the person who encounters a consequential mistranslation. If that person has no accessible mechanism for challenging the output, obtaining human review, correcting the record, reporting a recurring linguistic failure or seeking remedy when the error causes harm, then the assurance architecture terminates at the institution instead of reaching the user. Redress belongs inside the assurance problem rather than beside it.

The appropriate mechanism varies by deployment context because the consequences of failure vary enormously, and an awkward error in general informational content is not equivalent to an error affecting healthcare, employment, education, legal rights, benefits or emergency communication. Assurance should therefore examine not only average performance but foreseeable failure, escalation pathways and the institution's response when confidence in automated signed output is no longer justified. This is also why a human in the loop is inadequate as a generic assurance phrase, since the operative questions are which human, with what qualifications, available to whom, at what point, with what authority and within what response time. For DeafBlind people and others whose access requirements may not be satisfied by visual signed output alone, the availability of an automated signing system cannot by itself establish that effective communication has been achieved.

Five questions, and our own answers

The following are questions we believe buyers should put to organisations making consequential claims about signed language AI. We include ourselves within the scope of that scrutiny, and our current answers appear below, because publishing require-

ments for everybody else while quietly exempting ourselves would reproduce the exact governance problem this article criticises.

NCG's own answers to the five questions

The questions this article asks buyers to put to suppliers, answered by NCG as of October 2026. Every answer falls short in some respect, and we publish them in that condition.

Question	NCG's current answer	Status
What has your governance stopped in the last twelve months?	Claims have been revised or removed, but no register lets anyone verify it	○ **Not met**
Who assessed the signed output, were they Deaf, and what protocol did they use?	No documented assessment protocol or named rater criteria	○ **Not met**
Where did the signing material come from, and can the people in it withdraw it?	Provenance can be identified; no comprehensive withdrawal mechanism	◐ **Partly met**
Who can withhold approval, and what happens when they do?	Consultation exists; no Deaf-led body holds decision rights	○ **Not met**
What happens when a Deaf user says the output was wrong?	No redress requirement defined in NCG's own framework	○ **Not met**

Source: NCG self-assessment, October 2026 — **Novara Consulting Group**

Figure 3. NCG's own answers, October 2026. Four of five fall short entirely; one is partly met. Novara Consulting Group

What has your governance stopped in the last twelve months?

We cannot answer this with a single auditable record. We maintain governance instruments across our research programmes and apply an evidence standard that has, in practice, caused claims to be revised or removed before publication. We do not maintain a consolidated non-conformity, corrective-action or adverse-decision register through which an independent party could verify those assertions and trace the evidence behind each decision. That is a deficiency in our own evidence architecture, and believing that governance has affected our work is not equivalent to demonstrating it. We intend to establish a formal mechanism for recording material evidence-driven reversals, rejected claims, methodological failures and corrective actions, and to report from it periodically.

Who assessed the signed output, were they Deaf, and what protocol did they use?

We do not produce signed language AI output and cannot give a product-level answer. Where we commission, review or publish signed content, we do not operate a documented linguistic assessment protocol with named qualification criteria for Deaf assessors, a defined scoring instrument, inter-rater procedures and published decision thresholds. We would therefore fail the broader version of this question today. The absence of a generally accepted public assurance protocol for signed language output is a sector-level gap and it is not an excuse for describing an internal procedure as though it were a validated instrument. If we develop or adopt such a procedure, its methodology, limitations and evidentiary status should be disclosed plainly.

Where did the signed language material come from, and can the people in it withdraw it?

For signed material we hold, provenance can be identified. We cannot demonstrate a comprehensive revocation mechanism through which a signer can require source material and applicable downstream derivatives to be identified and withdrawn according to predetermined rules. That distinction matters, because knowing where footage came from and the person represented in it retaining meaningful control over its continuing use are not the same proposition. Revocation becomes particularly difficult once source material has entered derived datasets, trained systems, segment libraries or downstream products. We regard this as one of the hardest unresolved governance problems in the field and we do not claim to have solved it.

Who can withhold approval, and what happens when they do?

At present no external Deaf-led body holds formal veto authority over our work in this area. We have consultation, we do not have authority, and we will not describe consultation as though it were authority. Establishing a structure in which a named Deaf-led body holds documented decision rights, including a defined ability to withhold approval within an agreed scope, is work we have not completed. Until it exists, our governance claims should be read accordingly.

What happens when a Deaf user says the output was wrong?

We do not operate a signed language AI system deployed directly to Deaf users and therefore maintain no product-level linguistic redress mechanism. More significantly,

we have not established a complete assurance requirement defining how a disputed signed language output should be preserved, independently reviewed, corrected, escalated and connected to remedial action. Having argued above that redress belongs inside the assurance problem, we should hold that requirement explicitly in our own framework rather than leave it implied by general accountability language, and at present we do not.

Five questions produce five answers that fall short in some respect. We publish them in that condition deliberately, because waiting until every answer reflected well on us would defeat the purpose. Assurance is useful precisely when it can generate an inconvenient result.

The certificate is real, and so is the gap

The emerging AI assurance infrastructure deserves to be taken seriously. ISO/IEC 42001 provides a genuine management-system framework, ISO/IEC 42006 strengthens the conformity-assessment architecture around certification, accredited certification is now actually available rather than merely projected, and governments are investing in the tools, competencies and professional structures intended to make AI assurance more reliable and more widely accessible. Organisations procuring AI should want that infrastructure to succeed.

Signed language technology exposes what it cannot establish by itself. A management-system certificate can show that an organisation governs its AI activities through a defined system. It cannot, merely by existing, show that a Deaf person understands a signed output, that a community regards a use as legitimate, that revocability is meaningful where the technical and contractual architecture does not provide for it, or that a Deaf person harmed by an incorrect output has any route to challenge what happened. Those are not defects in the certificate. They are questions lying outside the proposition it was designed to prove, and that distinction needs to become ordinary procurement knowledge before signed language AI arrives routinely accompanied by the badges of AI assurance. Otherwise the market will produce a particularly dangerous form of confidence, which is confidence supported by legitimate evidence that has been attached to the wrong claim. What signed language AI still lacks is a domain-specific assurance layer capable of testing the proposition that matters at the point of use, which is whether Deaf people can understand, rely upon and challenge what the

system produces. Until that instrument exists the question remains open, and a certificate should not be allowed to close it.

NCG Insights publishes analysis on Deaf education, language access and the evidence base underpinning both.

Zitierweise

Grizzle, H. M. (2026, October 24). The Certificate That Says Nothing. Novara Consulting Group. https://www.novaracg.com/de/2026/10/24/the-certificate-that-says-nothing/