This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2026/11/14/who-checks-the-signing-ai-assurance-is-growing-up-around-sign-language-ai/.

Who Checks the Signing? AI Assurance Is Growing Up Around Sign Language AI

Heather Grizzle · November 14, 2026

AI assurance is booming around sign language AI, but it checks process, not signing. What Deaf-led evidence would look like, and what to ask.

Spis treści

1. What AI assurance means, in plain English
2. The assurance boom
3. Before assurance, there were warnings
4. Sign language AI is getting its own assurance layer
5. The gap: lots of paper, very little evidence
6. A benchmark is not a Deaf person
7. The research has been asking for Deaf evaluators for years
8. The signer nobody tested
9. Consultation is not assurance
10. What real assurance would look like
11. What you can ask, starting now
12. Where this is heading
13. Bibliografia
14. Tips

Deaf people have always run their own quality control. Put a new interpreter in front of a room of Deaf adults and the room knows within a minute whether that person can actually sign. Nobody hands out a form and nobody waits for a report. People

watch the hands, but mostly they watch the face, because that is where the grammar lives. Linguists describe signs by their parameters: handshape, movement, location and palm orientation, and then the fifth one, the non-manual markers, the eyebrows and mouth and shoulders and gaze that turn a statement into a question, mark a conditional, or tell you whether something happened once or kept happening. A hearing observer might watch an interpreter's hands and think everything is fine. The Deaf room watches the face and knows within a sentence or two whether the meaning is coming through or falling apart.

That instinct, checking the work before you trust it, is older than any technology standard. It is how Deaf communities have protected themselves from bad interpreters, bad teachers, bad captioning and bad policy for generations. When an interpreter was not working, the room said so, and if the room had any power, the interpreter was replaced. When the room had no power, which was most of the time in hospitals, courts and schools, people went home not knowing what had been said about their own bodies, their own cases, their own children.

Now machines sign. An AI avatar can render a train delay announcement into British Sign Language on a station screen. A phone can turn American Sign Language into English text. A kiosk in a hospital waiting room can sign the check-in instructions. A website can sign its terms of service. In most of these settings there is no Deaf room. There is one Deaf person, alone, in front of a screen, deciding whether to trust what it says, often with no way to ask a follow-up question and no way to tell anyone if it got something wrong.

The technology world has a name for the job the Deaf room used to do. It is called AI assurance, and it is one of the fastest-growing corners of the AI economy. Governments are funding it, standards bodies are writing rules for it, consulting firms are selling it, and regulators are starting to require it. Over the past eighteen months it has also started to wrap itself around sign language AI, through ethical frameworks, procurement toolkits, company quality policies, advisory committees and academic risk models.

This piece looks at that growth closely and asks one question of all of it: is any of this putting someone in the position the Deaf room has always been in, watching the actual signing and deciding whether it holds up? The short answer is not yet, mostly. The longer answer is that the pieces are being built, some of them by Deaf people, and the next two years will decide whether sign language AI assurance becomes a real check

on what these systems do or a stack of paperwork that makes buyers feel safe while users carry the risk.

W SKRÓCIE

- AI assurance is a fast-growing industry, and sign language AI now has its own layer of frameworks, toolkits, policies and advisory committees.
- Almost all of it is procedural assurance: proof that a process exists. Very little of it is evidentiary assurance: proof that Deaf people understand the signing.
- No sign language AI vendor in NCG's review of public evidence had published an independent evaluation of its signed output.
- The methods for Deaf-led output testing already exist in the research literature. What is missing is the requirement to use them.
- Anyone meeting sign language AI at school, work, a clinic or a transit system can ask eight assurance questions, listed below.

What AI assurance means, in plain English

AI assurance is the work of checking whether an AI system does what its makers say it does, and whether it is safe and fair to use for the people it touches. The term sounds technical, but the idea is ordinary. A restaurant passes a health inspection. A bridge gets load tested. An interpreter sits for certification. A new medicine goes through trials. In each case, somebody other than the seller looks at the thing and says whether it can be trusted, and what for.

The UK government, which has done more than any other to define this field, describes assurance as something that can be carried out by three kinds of actors: the companies that develop AI systems, the organizations that deploy them, and independent third-party assurance providers [7]. That three-way split matters. When a company assures its own product, you are getting the seller's view of the seller's work. When a hospital or a transit authority assures a product it bought, you are getting the buyer's view, which is better, but the buyer often lacks the expertise to judge sign language output and has every reason to believe the money was well spent. When an independent third party does it, you have a chance at an honest answer, but only if that third party is qualified, is not paid based on the result, and is free to publish what it finds.

There is a second split that matters even more for sign language, and it runs across all three kinds of actors. I call it procedural assurance versus evidentiary assurance. Procedural assurance proves a company has a process: there is a policy, a risk register, a named person responsible, a review step before release, a channel for complaints. Evidentiary assurance proves the output works: when real people use the system on real content, here is how often it was right, here is how badly it failed when it was wrong, and here is who was harmed. Think of the difference between a driving school proving it has a curriculum and a driver passing a road test. Both are useful. Only one of them tells you whether the person behind the wheel can drive.

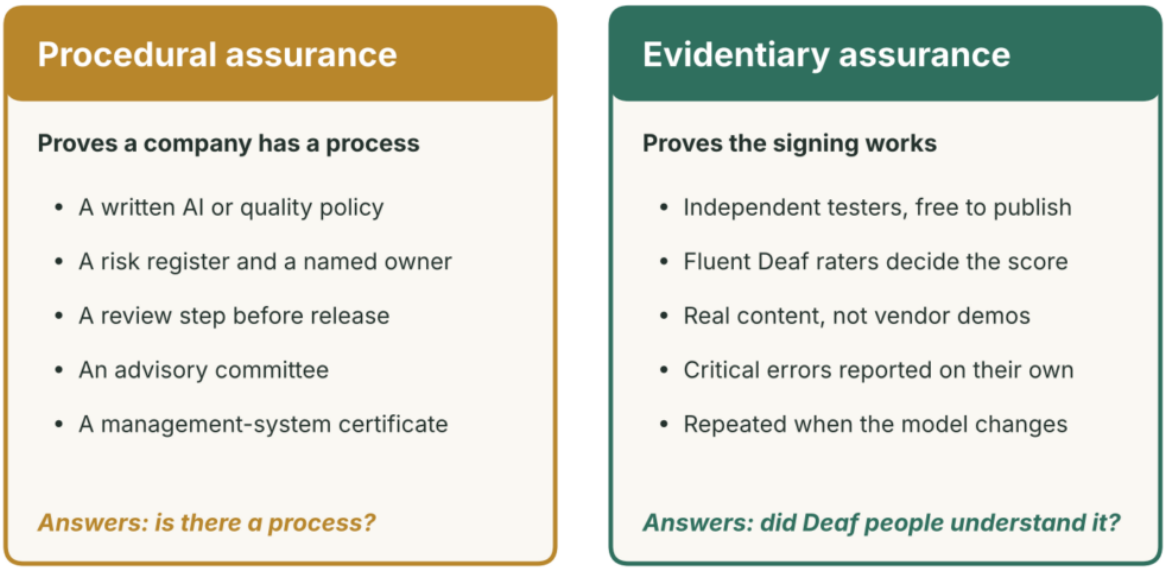


Figure 1. Two kinds of assurance. Procedural assurance answers whether a process exists. Evidentiary assurance answers whether Deaf people understood the signing.

A company can be fully assured on process and still ship an avatar that drops the facial grammar that turns a statement into a question. That is not a hypothetical worry. It is the central weakness of almost every sign language AI system on the market, and process paperwork cannot see it, because process paperwork does not watch the face.

The best-known example of procedural assurance is ISO/IEC 42001, the international standard for AI management systems, published in December 2023 [12]. It is a real and useful standard. It asks organizations to set AI policies, assess risks, assign responsibility, monitor systems and keep improving them, and organizations can be cer-

tified against it by an external auditor. But a certificate under ISO 42001 covers how an organization manages AI within a stated scope. It is not a statement that any particular AI output is accurate, and anyone describing a product as "ISO 42001 certified AI" is overstating what an auditor looked at [3].

In July 2025 the standards bodies added ISO/IEC 42006, which sets the requirements for the bodies that audit and certify organizations against ISO 42001 [13]. In plain English, it audits the auditors: it sets rules for their competence, impartiality and methods. That is a sign of a maturing field, and it should make 42001 certificates more trustworthy. But the accounting profession's own commentary was careful to point out that 42006 applies to one specific kind of audit, the audit of a management system, and is not a standard for every kind of AI audit or assurance activity [10]. It still does not put a fluent signer in front of the output.

So when you hear that a sign language AI company is "assured," "audited," "certified" or "governed," the first question is always the same. Assured on what? On its process, or on its signing?

The assurance boom

To understand why assurance is arriving in sign language AI now, it helps to see how much money and policy energy is moving into assurance generally.

The United Kingdom has measured its own assurance market more carefully than anyone else. In November 2024 the Department for Science, Innovation and Technology published its first analysis of the field, estimating that 524 firms were supplying AI assurance goods and services in the UK, generating about 1.01 billion pounds in gross value added and employing roughly 12,500 people [6]. Only 84 of those firms were specialised assurance companies. The rest were AI developers assuring their own products, diversified firms like consultancies and accountancies adding assurance to their menu, and organizations doing it in-house. The government said the market could grow six-fold within a decade if barriers were addressed.

THE UK AI ASSURANCE MARKET

524
firms supplying AI assurance in the UK in 2024

84
of them specialised assurance companies

£1.01bn
gross value added in 2024

£18.8bn
projected market by 2035

Source: UK Department for Science, Innovation and Technology, 2024 and 2025

Ten months later, in September 2025, the same department published its Trusted Third-Party AI Assurance Roadmap, focused on the independent providers who check AI for organizations that cannot do it themselves [7]. The roadmap named the problems plainly: providers struggle to find qualified people, face skills shortages, cannot always get the information they need from the developers whose systems they are checking, and are trying to keep pace with technology that changes every few months. It weighed different ways to improve quality, including professional certification for

individual assurers and certification of assurance processes, and committed the government to work on a skills and competencies framework for the profession. Its headline projection was that the UK assurance market could reach 18.8 billion pounds by 2035 [2]. One commentator noted a gap that matters a great deal for public services: the roadmap discusses public sector use but makes no commitment to require third-party assurance when government buys AI [1]. A year later, in September 2026, the UK tech industry gathered in London to take stock of what the roadmap had produced, with regulators, investors and insurers in the room alongside the assurance firms themselves [28].

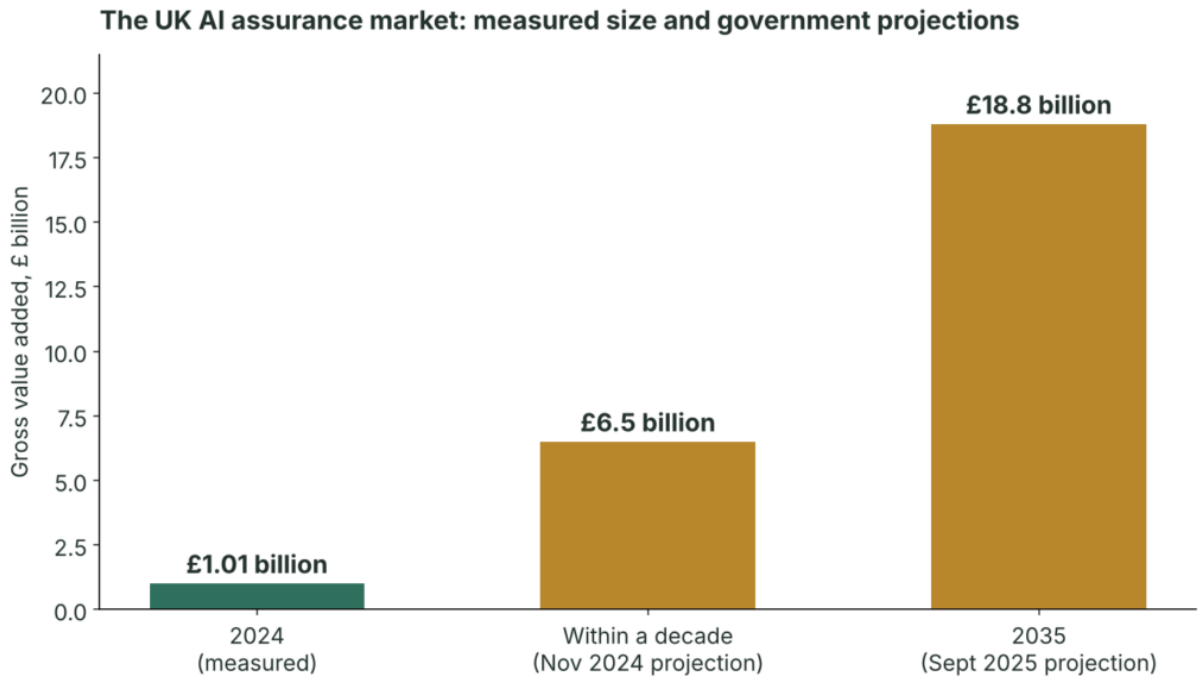


Figure 2. The UK AI assurance market: measured size and two government projections.
Sources: DSIT, Assuring a Responsible Future for AI (2024); DSIT, Trusted Third-Party AI Assurance Roadmap (2025)

That last detail is worth sitting with. Investors and insurers care about assurance because it is how they price risk. When an insurer will not cover an AI deployment without evidence that it works, or covers it at a lower premium with evidence, assurance stops being a nice-to-have and becomes a cost of doing business. That is the moment a market forms.

In Europe, the driver is regulation. The EU AI Act entered into force in August 2024 with most of its heavier obligations scheduled for August 2026. This summer the EU amended the Act through what it calls the Digital Omnibus on AI, Regulation (EU) 2026/1744, which entered into force on July 27, 2026 [22]. The amendment pushed back the obligations for high-risk systems listed in Annex III of the Act, the category that covers uses like education, employment and access to essential services, to December 2, 2027, and the obligations for high-risk AI embedded in regulated products to August 2, 2028. Commentators were quick to warn that this was a delay, not a pause, and that most of the Act's transparency obligations stayed on track [19].

One of those near-term obligations lands very soon. From December 2, 2026, providers of generative AI systems must mark AI-generated content in a machine-readable way so it can be detected as synthetic [15]. For companies producing synthetic signers, digital twins of human signers, or AI-generated signed video, that date is less than three weeks from when I am writing this. How each vendor reads that obligation for its own product is a legal question I cannot answer for them, but it is a question every buyer in Europe should be asking.

There is also a deeper reason sign language AI will keep running into the AI Act. Researchers presenting at the International Telecommunication Union this July pointed out that sign language communication requires facial expression, gaze and body movement, and that those elements are biometric data, which the Act treats as high-risk in many deployment contexts [23]. Every sign language system that watches a signer is, in effect, processing a face. That puts the field closer to the Act's most demanding rules than many vendors would like to admit.

Money follows regulation, and assurance follows money. When buyers know they will eventually have to show a regulator, an insurer or a court evidence that the system they bought was fit for purpose, they start paying for that evidence. That is why the assurance market is growing, and it is why sign language AI is being pulled into it.

Before assurance, there were warnings

Deaf organizations saw this coming long before anyone called it assurance. In March 2018 the World Federation of the Deaf and the World Association of Sign Language Interpreters issued a joint statement on signing avatars [29]. It was not a rejection of the technology. It acknowledged that avatar image quality had improved and that the

technology had real potential. Its concern was narrower and sharper: the way public authorities were deciding where and when to use avatars as a form of access. Because signed languages are full languages with their own structures, the statement explained, word-for-word translation from a spoken language is not possible, and any translation has to account for context and culture. On that basis the two organizations cautioned against using signing avatars as a replacement for human signers.

The statement singled out the situations where the risk was highest: information that is live, complex, or of significant importance to the lives of Deaf citizens. It also made two technical points that still hold up. Digital sign language dictionaries did not yet contain a full corpus of signs and their variants, so the raw material for generating good signing did not exist. And avatars were a one-way channel, from spoken or written language into sign, with nothing going back the other way.

Read that statement now and you can see the shape of an assurance framework hiding inside it. It sorts uses by risk. It names the conditions under which the technology is not ready. It insists that the quality question be judged against the language as Deaf people actually use it, including regional and social variation, rather than a simplified standard version. It asks that sign language work be led by Deaf native signers. What it did not have, in 2018, was an industry of assurance providers, a regulator asking for evidence, or buyers willing to pay for it. Eight years later, some of those things finally exist. The question is whether they have absorbed what the Deaf organizations said first.

HOW SIGN LANGUAGE AI ASSURANCE HAS DEVELOPED

March 2018

WFD and WASLI statement on signing avatars

Cautions against avatars replacing human signers, especially for live, complex or high-stakes information.

September 2023

Interpreting SAFE-AI Task Force launches

Interpreters, language companies, health systems, state governments and Deaf community members organize around AI in interpreting.

2024

SAFE-AI guidance and the Deaf-led research review

Ethical guidance for AI in interpreting; a Deaf-led review of 101 sign language AI papers finds the field lacks meaningful Deaf input.

2025

EUD framework, CoSET toolkit, BSL procurement report

Fifteen ethical principles; buyer checklists; a call for a Deaf-majority oversight body in the UK.

2026

Vendor policies and advisory committees

Signapse publishes a quality policy; Google launches SL2T with an advisory committee; ITU researchers publish a risk workflow.

Grudzień 2026

EU marking of AI-generated content

Providers of generative AI must mark synthetic content in a machine-readable way.

December 2027

EU high-risk obligations apply

Annex III high-risk rules take effect after the Digital Omnibus delay.

Sign language AI is getting its own assurance layer

Over the past eighteen months, a recognizable assurance layer has formed around sign language AI specifically. It comes from four directions at once: Deaf organizations, the interpreting profession, researchers, and the companies themselves. Each piece is worth understanding on its own terms, because each one is good at something different and each one leaves a different gap.

From Deaf organizations

The European Union of the Deaf published an Ethical Framework on AI and Sign Language, developed with Prof. Dr Filipe Venade, alongside a longer book on sign language in the era of AI [9]. The framework sets out fifteen principles for developers and researchers, covering co-creation with Deaf communities, cultural respect, consent for the use of signed recordings, and data protection. Two of its principles go straight to assurance. One says that using sign language data beyond its original intent requires renewed consent. Another says developers must not alter sign language recordings in

ways that distort or misrepresent them, and that they are accountable for harm caused by misuse or misrepresentation of sign language data. The framework is now listed in the OECD's catalogue of tools for trustworthy AI [21], which means it sits in the same reference library governments and companies use when they look for governance tools. That is a quiet but real achievement: a Deaf-authored standard is now part of the mainstream assurance toolkit.

From the interpreting profession

In the United States, the most developed work has come out of the Interpreting SAFE-AI Task Force, a coalition of interpreters, language companies, healthcare providers, state governments and Deaf community members that launched in September 2023 [24]. In June 2024 it published its Guidance for the Safe and Ethical Use of AI in Interpreting, built on a perception survey run by CSA Research and a qualitative study by its Advisory Group on AI and Sign Language Interpreting [25]. That advisory group, which later became the Coalition for Sign Language Equity in Technology (CoSET), coined the slogan #DeafSafeAI and had fourteen members on the final review team for the guidance. Its principles include end-user autonomy, safety and wellbeing, transparent operations, human oversight, and privacy and security.

In August 2025 the Task Force and CoSET released Part A of their AI Interpreting Solutions Evaluation Toolkit, the most practical assurance tool in the field so far [26]. It is written for the people who make buying decisions: language access directors, compliance and legal teams, accessibility offices, technology leaders and procurement staff. It contains five sets of checklists. The first covers organizational readiness across eight areas, from strategy and governance to quality assurance and rollout. The second gives setting-specific guidance for healthcare, legal, education and business. The third is a risk framework for scoring each use case and defining when it must escalate to a human interpreter. The fourth assesses vendors across ten categories, including accessibility, technical fitness, privacy, ethics, escalation and compliance. The fifth gives template RFP language that requires pilots and human backup. The coalition presented its risk framework at the SLxAI Summit in April 2026 [18]. Two further parts are planned: Part B, on technical specifications, is supposed to set out baseline model expectations and performance metrics, and Part C will translate the legal environment into procurement criteria and policy language. Part B is the piece that would move the toolkit from asking good questions toward measuring answers.



From researchers

In the UK, researchers at the Minderoo Centre for Technology and Democracy and Newcastle University published a report through the British Deaf Association in September 2025 on how British Sign Language AI is being bought by public services [17]. Their central finding was that BSL AI is being procured without adequate Deaf involvement, and that procurement often treats BSL as a technical accessibility problem rather than a complete language with legal standing under the BSL Act 2022. They named five linked harms: communication breakdown from inaccurate output, linguistic erasure as BSL gets reduced to simplified gestures, developmental harm to Deaf children if AI becomes their first exposure to BSL, displacement of qualified interpreters and translators, and loss of trust between Deaf communities and public bodies. They also drew a distinction every buyer should learn: translation, where output can be reviewed before release, carries less risk than real-time interpreting, where errors cannot easily be caught or corrected.

Their recommendations are the most concrete assurance proposals I have seen in this field. They call for an independent BSL AI oversight body with a Deaf majority and real powers to monitor, evaluate deployed systems, handle complaints, investigate and publish findings. They want contract performance indicators that measure linguistic accuracy, user satisfaction and community trust, not just technical uptime. They want public bodies to publish annual transparency reports on their BSL AI under the UK's Algorithmic Transparency Recording Standard, in BSL, open to community review. They want algorithmic impact assessments co-designed with Deaf experts, starting with a needs-first question about whether AI is the right solution at all. And they want Deaf communities to have an explicit right to recommend against deployment. Their most striking observation is that the main barriers to doing this well are procedural, not technical or financial. Nobody is short of the money or the engineering to assure BSL AI properly. They are short of the process that would require it.

Internationally, researchers presenting at the ITU's Kaleidoscope conference in July 2026 laid out a full risk management workflow for AI signing avatars, starting with identifying biometric and linguistic data at the point of collection, validating annotation quality, and ending with validation by Deaf stakeholders

> The barriers to doing BSL AI well are procedural, not technical or financial.

and accessibility experts [23]. It is the closest thing yet to an engineering lifecycle for sign language AI assurance.

> Minderoo Centre for Technology and Democracy, paraphrased

From the companies

Vendors have moved too, and they deserve credit for publishing what they have. Signapse now posts an AI governance framework on its website, with principles including that sign languages are full languages, that people should control how their signing and digital likeness are used, and that people should know when AI is being used [27]. Alongside it sits a Translation Quality Assurance Policy, signed by the CEO and dated February 2026. It lays out a seven-step workflow: identify the content, audience, sector and risk level; decide whether AI translation is appropriate at all; generate the translation; review it for accuracy, clarity and cultural meaning; deliver it with AI output labeled where appropriate; collect feedback; and use that feedback to improve. It names who is responsible for quality, lists the metrics the company tracks, and commits to reviewing the policy every six months.

Google DeepMind, when it launched its SL2T sign-to-text model on the Pixel 11 in August 2026, created an AI Sign Language Advisory Committee of global Deaf organizations and subject-matter experts to guide responsible deployment [4]. According to Google, that committee co-authored a joint impact report on the model's capabilities and limitations.

Set side by side, this is real growth. Five years ago almost none of it existed. Deaf organizations have written standards that now sit in international catalogues. The interpreting field has built buyer checklists. Researchers have mapped harms and proposed oversight bodies. Two of the most visible companies in the field have published governance material. If you were measuring assurance by the volume of documents, sign language AI would look like one of the better-governed corners of AI.

The gap: lots of paper, very little evidence

Here is the problem. Read all of that material closely and ask one question of every piece: who, outside the company that built the system, has checked the signing?



	Sets principles or process	Gives buyers questions	Independent check of the signing	Published results scored by Deaf raters
EUD Ethical Framework (2025)	Yes	Partly	No	No
SAFE AI / CoSET Toolkit, Part A (2025)	Yes	Yes	Planned (Part B)	No
Minderoo / BDA BSL report (2025)	Yes	Yes	Proposed	Proposed
Signapse quality policy (2026)	Yes	No	No	No
Google SL2T launch (2026)	Partly	No	No	No
ISO/IEC 42001 certificate	Yes	No	No	No

Based on each source's own public documents as of November 2026. 'Proposed' and 'Planned' mean written down, not yet operating.

Figure 3. What each piece of the sign language AI assurance layer actually checks, based on each source's own public documents as of November 2026.

Start with the company documents, because they make the most direct claims about quality. Signapse's quality policy is more detailed than most, and it describes a real review process. But the checks it describes are internal: review by the company's own sign language specialists, internal audits on a regular schedule, and feedback from Deaf user testing groups, consultation with Deaf organizations, and real projects and trials. It talks about future external reviews with the Deaf community. It lists AI translation accuracy targets among its metrics, but gives no figures for them. It does not describe any independent third-party evaluation of its output, and it does not mention any certification. The laws it lists, including the EU AI Act and the European Accessibility Act, are laws it says it follows, not audits it has passed.

Google's SL2T announcement is careful and technically impressive. It reports a score of 70 on BLEURT, an automatic translation quality metric, against the FLEURS-ASL test set, and describes that as significantly higher than any previously reported score [4]. It shows example translations and openly notes the error types that remain, including rare signs, fingerspelling, classifiers and tense. It mentions Deaf user studies and an impact assessment with Deaf experts. What the public announcement does not describe is independent evaluation of the model's output, or panels of Deaf raters scoring its translations, and the advisory committee's individual members are not named. The joint impact report may contain more, and I would welcome it if it does. But

the evidence a Deaf user can see from outside is a benchmark number and a description of a process.

The SAFE AI and CoSET toolkit is the strongest buyer tool in the field, and it is honest about its own stage of development. Part A teaches buyers what to ask. Part B, the piece with performance metrics, is still to come. A checklist is only as good as the evidence that comes back when a buyer fills it in, and right now there is very little independent evidence for the vendor to hand over.

None of this is a scandal. Writing a policy, forming a committee and publishing a benchmark are legitimate first steps, and plenty of companies in this market have done far less. But all of it is procedural assurance. It tells you a process exists. It does not tell you whether a Deaf patient reading an AI signer on a hospital screen understood the instruction correctly, or whether a Deaf commuter on a station platform understood which train had been cancelled.

My firm has spent much of this year reviewing the public evidence for the sign language AI vendors selling into English-language markets, using a structured evaluation framework I built for that purpose. The clearest pattern across the roster was this: none had published an independent evaluation of its signed output. The pattern underneath it was just as important. The documentation did not track the stakes. The largest claims, about serving public services, standing in for interpreters and scaling access across whole sectors, tended to rest on the thinnest public evidence. And the routes for a Deaf user to report a problem and have it reach someone accountable were rare, and the one I located had been set up by a buyer for a pilot, not by the vendor. Those patterns are what the whole assurance conversation has to close, and none of them can be closed by writing another policy.

A benchmark is not a Deaf person

It is worth slowing down on benchmarks, because they are the form of evidence the industry produces most readily, and they are easy to mistake for assurance.

A benchmark is a fixed test set with an automatic scoring method. For sign-to-text systems, you take a collection of signed videos with agreed English translations, run the model on them, and use a metric to measure how close the machine's English is to the reference English. BLEURT, the metric Google used, is a learned metric that compares meaning rather than just counting matching words, which makes it better than



older measures. A high score genuinely means the model is getting closer to the reference translations on that test set. That is real progress and it should be reported.

But notice what a benchmark cannot do. It cannot tell you how the system performs on signers who look nothing like the people in the test set: older signers, Black ASL users, signers with regional variation, children, signers with motor disabilities, signers filmed in poor light. It cannot tell you which errors were harmless and which ones changed the meaning of a medical instruction. It averages everything into one number, so a system that is excellent on everyday phrases and dangerous on numbers, negation or dates can still post a strong score. And it says nothing at all about the reverse direction, text into sign, where the output is a signer that a Deaf person has to watch and understand. For that direction, the only valid test is a fluent Deaf viewer.

This is not a new insight. Research on Sign Language of the Netherlands avatars in 2021 deliberately measured three separate things: whether Deaf users recognized the individual signs, whether they understood the sentences as intended, and how clear the translations were [20]. Those are different questions with different answers. A viewer can recognize every sign in a sentence and still not understand what the sentence means, because the grammar that holds the signs together, much of it on the face and in the use of space, is where machines most often fail. A benchmark that scores signs or words cannot see that failure. A Deaf viewer sees it immediately.

The research has been asking for Deaf evaluators for years

If you want to know why sign language AI assurance is weak on evidence, the academic literature has already given a large part of the answer.

In 2024 a team of Deaf and hearing researchers, including Maartje De Meulder and Julie Hochgesang, published a systematic review of 101 recent papers in sign language AI [5]. They found a field dominated by hearing researchers who do not sign, with a tendency to chase perceived communication barriers rather than problems Deaf people actually named, to train on unrepresentative datasets, to rely on annotations without a sound linguistic foundation, and to build on flawed models. Their position was that the field lacks meaningful input from Deaf stakeholders and is instead shaped by whatever decisions are most convenient or seem most important to hearing researchers. Their call to action was that Deaf researchers must lead.



That same year, a user study presented at a major computational linguistics conference said it plainly: research has shown a significant deficiency in evaluating sign language generation systems with the people who are deaf and use sign language [16]. Researchers in Kazakhstan, testing avatars for Kazakh-Russian Sign Language over several versions, made the involvement of Deaf signers central to their evaluation design and found that Deaf signers still preferred human agents and were less optimistic about avatars than about interpreters, while remaining open to the idea for future use [14]. Earlier work on Irish Sign Language avatars, run with Deaf participants at a Deaf village in Ireland, found something every assurance designer should notice: participants tended to underestimate their own comprehension when asked directly, which means simply asking Deaf users how well they understood is not a reliable measure on its own [11]. You need comprehension questions, not just satisfaction ratings. And the EU's EASIER project ran its end-user evaluations of a signing avatar with Deaf groups in their own local sign languages, with anonymity protected, which is the kind of design detail that separates a real evaluation from a demo [8].

The methods for evidentiary assurance in sign language exist. They have been tested, refined and published for more than a decade. What has been missing is not knowledge. It is the requirement to use them, the money to pay for them, and an independent party with the standing to publish the results whether the vendor likes them or not.

The signer nobody tested

There is a second gap, and it is personal for me. I am DeafBlind. Almost every sign language AI product on the market makes the same quiet assumption: that the person on the other end can see a screen at an ordinary distance, in ordinary light, with a full visual field, at whatever speed and size the designer chose. An avatar signing on a station display mounted above head height, a kiosk with glare on the glass, a phone app that renders a small signer in the corner of the screen, a sign-to-text system that needs the user to frame themselves neatly in a camera: each of these works fine in the lab and fails people like me in the world.

It goes further than screens. Many DeafBlind people communicate through tactile signing or Protactile, where language travels through touch rather than sight. No camera-based system can read it, and no avatar can produce it. That is not a criticism of any one company. It is a fact about the technology. But it becomes an assurance pro-

blem the moment a public body buys a sign language AI product and describes the result as sign language access, because a whole group of signers has been excluded and nobody wrote that down.

Assurance processes that never test with DeafBlind, low-vision, older or disabled signers will certify systems that leave those people out, and nobody will notice, because nobody on the review was in a position to notice. The same is true of signers whose variety of a sign language is underrepresented in training data, which is the problem the World Federation of the Deaf was pointing at in 2018 when it said sign language work should reflect all the different signs used by a language community. An assurance result is only valid for the people it tested. If the test panel was young, sighted signers from one city, the result applies to young, sighted signers from one city. Anyone selling it as more than that is overstating the evidence.

Consultation is not assurance

One trap deserves its own section, because Deaf community members are the ones most likely to be pulled into it, and many readers of this piece will be asked to take part in it sooner or later.

As sign language AI companies come under more scrutiny, they are inviting more Deaf people in. They form advisory boards and committees. They run Deaf user testing groups. They hire Deaf staff and consultants. They feature Deaf signers in launch videos and Deaf leaders in press releases. Some of this is sincere and some of it produces better products. But all of it carries a risk I have written about before as the Deaf halo effect: the presence of Deaf people around a product gets read as evidence that the product works, or that it has been governed, when no Deaf person has actually been given the authority to test it, judge it and say no.

Consultation and assurance are different things. Consultation means the company asked for your views. Assurance means someone with the right expertise tested the output against a standard, and the result can stop or change the deployment. A Deaf advisory committee whose members are not named, whose advice is not published, and whose power to block a launch is not stated is consultation. It may be good consultation. It is not assurance, and it should not be described as if it were.

The Minderoo researchers drew this line carefully in their BSL report. They did not just ask for Deaf advisory groups. They asked for permanent Deaf advisory groups

with decision-making authority, a Deaf-majority oversight body with power to investigate and publish, and an explicit right for Deaf communities to recommend against using AI at all [17]. Those three features are what turn participation into accountability: authority over the decision, independence in what you can say publicly, and the option to say no.

If you are asked to advise a sign language AI company, join its testing group, or appear in its materials, it is entirely fair to ask a few questions before you agree. What can my input change? Will my findings be published, and can I publish them myself if the company disagrees? Am I being paid, and is my payment tied in any way to what I conclude? Will my name, face or signing be used in marketing, and can I withdraw it later? Who else is on this committee, and will their names be public? Will DeafBlind, low-vision, older and regionally diverse signers be included, or is it just me? The answers will tell you whether you are being asked to do assurance or to decorate it. Both answers are useful to know. Only one of them should be described to buyers as evidence.

Interpreters face a version of the same trap. When a vendor says its product was built with interpreters or reviewed by interpreters, that is worth something, but interpreters are not the end users of sign language AI, and their judgment of whether the signing is acceptable is not a substitute for Deaf people's judgment of whether they understood it. The strongest designs put Deaf raters in the deciding seat, with interpreters and Deaf interpreters contributing their specific expertise alongside them.

What real assurance would look like

Evidentiary assurance for sign language AI is not a mystery. The pieces already exist, in the research literature, in other fields of AI, and in the way interpreter certification has worked for decades. What follows is not a product spec. It is the set of conditions that, taken together, separate a real check on the signing from a document about a process.

It has to be independent

The people testing the system should not be paid more or less depending on the result, should not be owned by or financially tied to the vendor, and should have an unconditional right to publish what they find. The vendor can have a right of reply,

which is fair. It should not have a right of approval. An evaluation the vendor can veto is marketing with extra steps.

Fluent Deaf raters have to decide

The people whose judgment sets the score should be fluent Deaf users of the language being tested, with interpreters and Deaf interpreters contributing their expertise alongside them. This is not a diversity requirement. It is a validity requirement. The thing being measured is whether Deaf people understand the output and whether the signing is grammatical, and the people best qualified to judge that are Deaf signers. A test without them is measuring something else.

It has to measure understanding, not just preference

As the Irish research showed, people are not always good judges of their own comprehension, so asking Deaf users whether they liked an avatar or felt they understood it is not enough. Good evaluations ask comprehension questions about the content, compare the AI against a human signer on the same material, and report how big the difference is.

It has to use real content, not demos

The test material should come from the setting where the system will actually be used: real station announcements, real hospital instructions, real classroom content. Clips the vendor picked, polished in advance or re-rendered after a failure should be ruled out. And some portion of the material should be deliberately hard, with numbers, dates, negation, names, conditionals and the kinds of grammar machines most often get wrong, because that is where harm happens.

It has to count harm, not average it away

Errors are not equal. A slightly awkward sign in a weather update is not the same as a dropped negation in a medication instruction. Evaluations should sort errors by severity and report critical errors, the ones that could change a medical decision, a legal outcome or someone′s safety, on their own line, never folded into a nice overall percentage. A system that is ninety-five percent accurate overall and wrong about which pill to stop taking has failed.

It has to match the stakes

A pre-recorded translation of a museum label, reviewed by a Deaf translator before release, needs a lighter level of evidence than a live system signing emergency announcements or interpreting a doctor's appointment in real time. The Minderoo report's distinction between translation and interpreting, and the SAFE AI toolkit's risk scoring, both point in this direction. The more harm an error could cause, and the less chance there is to catch it before a Deaf person relies on it, the more evidence should be required before deployment.

It has to cover both directions and every kind of signer

Systems that produce signing need to be judged on whether Deaf viewers understand them. Systems that read signing need to be judged on how often they misread, miss or quietly drop what was signed, and those error rates need to be broken down by signer and setting: skin tone, lighting, camera angle, signing style, age, regional variety, one-handed signing, and disability. A recognition system that works well for some signers and badly for others will look fine in an average.

It has to be repeated

Models change constantly. A result on last year's version says nothing about this year's. Assurance results should be tied to a specific version, expire after a set period, and be redone when the system changes in ways that could affect quality.

It has to keep going after launch

The best pre-deployment test still cannot predict everything that happens in the field. Real assurance includes a working channel for Deaf users to report problems, reachable in sign language, that leads to someone accountable, with reports reviewed by someone other than the vendor and summarized publicly. The Minderoo proposal for annual transparency reports in BSL is a good model for this.

None of these conditions is exotic. Every one of them is standard practice somewhere: in clinical trials, in aviation safety, in interpreter certification, in the better parts of the general AI assurance market. What sign language AI lacks is not the method. It is the expectation that the method will be used, and buyers who refuse to deploy without it.

The Minderoo report's Deaf-majority oversight body is the institutional version of the same idea. Whether the check comes from a regulator, a community body, an inde-



pendent evaluator or some combination, the principle is the one the Deaf room has always run on: the people who rely on the signing are the ones who check it.

What you can ask, starting now

Most of the people reading this will never run an evaluation or sit on a procurement panel. But almost everyone reading this will, sooner or later, meet sign language AI at a school, a job, a doctor's office, a transit system, a bank or a government agency. When that happens, you are the Deaf room. You do not need technical training to ask assurance questions, and institutions that bought these systems should be able to answer them. These eight will tell you most of what you need to know.

1. Has anyone outside the company tested this system's signing, and can I read the results? This is the most important question, and the answer separates evidence from process. A vendor's own testing, a benchmark score or a statement that the product was developed with the Deaf community is not the same as an independent evaluation. If there is no independent result to read, the system has not been assured on its output.
2. Who scored the output? Ask whether fluent Deaf signers were the ones judging the signing, how many of them, and whether DeafBlind, low-vision, older and regionally diverse signers were included. If the answer is hearing staff, interpreters alone, or nobody, the result does not tell you what you need to know.
3. Was it tested on content like what I will see here? A test on weather updates says little about medical instructions. Ask whether the evaluation used material from this setting, and whether it included the hard parts: numbers, dates, names, negation.
4. What happens when it gets something wrong? Ask how you reach a qualified human interpreter, how fast, and whether you can do it in the moment rather than after the fact. The SAFE AI toolkit treats human backup as a basic procurement requirement. If there is no human route, that is a decision someone made, and you are entitled to know who.
5. Is the AI-generated signing clearly labeled as AI? You should never have to guess whether you are watching a person or a machine. In Europe, machine-readable

marking of AI-generated content becomes a legal obligation for providers on December 2, 2026. Clear visible labeling is simply good practice everywhere.

6. What data is being collected from me? If the system watches you sign, it is capturing your face, your body and your movements. Ask whether your video is stored, whether it is used to train future models, and whether you can say no without losing access. The EUD framework treats reuse of signed recordings beyond their original purpose as something that needs fresh consent.
7. Who do I report an error to, and does anyone outside the vendor see those reports? A complaint that goes only to the company that built the system is customer service, not oversight. Ask whether reports are reviewed independently and whether a summary is published.
8. Was there a choice not to use AI here? The Minderoo researchers argue that every assessment should start by asking whether AI is the right solution at all, and that Deaf communities should be able to recommend against it. Ask whether that question was asked, and who answered it.

If the answer to the first question is no, the system has not been assured. It has been described. That is not necessarily a reason to refuse to use it, especially for low-stakes content where any access is better than none. But it is a reason to treat its output with the same caution the Deaf room has always applied to an interpreter nobody has seen sign before, and it is a reason to put the question to the institution in writing. Institutions respond to questions they receive in writing, and every written question is a small piece of demand for real assurance.

Where this is heading

AI assurance in sign language AI is growing, and that is good news. The frameworks, toolkits, policies and committees of the last eighteen months are the scaffolding a real system of accountability will need. But scaffolding is not a building, and the next two years will decide what gets built on it.

There are two ways this can go. In the first, assurance in sign language AI grows the way it has grown in many other parts of the AI market: as paperwork. Vendors obtain management-system certificates, publish principles, form advisory committees and post benchmark scores. Buyers collect those documents, check the boxes on a procu-

rement form and feel protected. Regulators, stretched thin and focused on bigger industries, accept the documents at face value. Deaf users keep meeting signing that no Deaf person independently checked, and when it fails them, there is a file somewhere showing that everyone followed the process.

In the second, assurance grows as evidence. The SAFE AI toolkit′s Part B sets real performance expectations. Buyers start writing independent, Deaf-led evaluation into their contracts as a condition of payment. Public bodies in the UK take up the Minderoo recommendations and publish transparency reports in BSL. Insurers, who price risk for a living, start asking for evaluation results before they cover deployments in healthcare and transport. Independent evaluators with Deaf raters at their core become a normal part of the market, the way independent testing labs are normal for electrical products. Vendors with good systems welcome it, because evidence is how a good product proves it is different from a bad one.

Which way it goes is not up to the vendors alone. It depends heavily on what buyers demand, and buyers respond to pressure from the people their services are supposed to reach. The European high-risk deadline of December 2027 gives everyone a window. The UK roadmap′s silence on requiring third-party assurance in public procurement is a gap that Deaf organizations can push to fill. And the interpreting field′s toolkit is a ready-made vehicle for raising the bar, if Part B is written with Deaf-led output testing at its center.

There is also an opportunity here that I want Deaf readers in particular to see. AI assurance is becoming a profession. The UK government is building a skills and competencies framework for it. Standards bodies are writing rules for who is qualified to audit. Firms are hiring. And in sign language AI, the single most important qualification for checking the output, fluent knowledge of a signed language and lived experience of whether signing makes sense, is one that Deaf people hold and most of the existing assurance industry does not. Learning how AI works, how it is evaluated and how it is governed is not a side interest for Deaf professionals. It is a way into the jobs that will decide whether this technology is safe for our communities.

The Deaf room has always known how to check the signing. The work now is making sure that room gets a seat, with authority, wherever these systems are bought, built and judged. Deaf communities are in the best position of anyone to define what counts as evidence in sign language AI. We should make sure we are the ones who do it, before someone else defines it for us.

Bibliografia

1. AI Governance Library. (2025). Trusted Third-Party AI Assurance Roadmap (UK Government, DSIT, September 2025) [Review]. https://www.aigl.blog/trusted-third-party-ai-assurance-roadmap-uk-government-dsit-september-2025/
2. Burges Salmon. (2025). Trusted third-party AI assurance: DSIT roadmap. https://www.burges-salmon.com/articles/102l5fo/trusted-third-party-ai-assurance-dsit-roadmap/
3. Copla. (n.d.). What is ISO 42001? https://copla.com/blog/compliance-regulations/what-is-iso-42001/
4. Google DeepMind Sign Language Team. (2026, August 12). Putting sign language AI into users' hands. Google DeepMind. https://deepmind.google/blog/putting-sign-language-ai-into-users-hands/
5. Desai, A., De Meulder, M., Hochgesang, J. A., Kocab, A., and Lu, A. X. (2024). Systemic biases in sign language AI research: A Deaf-led call to reevaluate research agendas. Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages. https://aclanthology.org/2024.signlang-1.6.pdf
6. Department for Science, Innovation and Technology. (2024, November 6). Assuring a responsible future for AI. GOV.UK. https://www.gov.uk/government/publications/assuring-a-responsible-future-for-ai/assuring-a-responsible-future-for-ai
7. Department for Science, Innovation and Technology. (2025, September 3). Trusted third-party AI assurance roadmap. GOV.UK. https://www.gov.uk/government/publications/trusted-third-party-ai-assurance-roadmap/trusted-third-party-ai-assurance-roadmap
8. The EASIER mobile application and avatar end-user evaluation. (2024). Proceedings of the LREC-COLING 2024 11th Workshop on the Representation and Processing of Sign Languages. https://www.sign-lang.uni-hamburg.de/lrec/pub/24021.pdf

9. European Union of the Deaf. (2025). Ethical framework on AI and sign language. https://eud.eu/wp-content/uploads/2025/07/ETHICAL-FRAMEWORK-ON-AI-AND-SIGN-LANGUAGE.pdf
10. ICAEW Insights. (2025, August 1). New standard for firms certifying AI management systems. https://www.icaew.com/insights/viewpoints-on-the-news/2025/aug-2025/new-standard-for-firms-certifying-ai-management-systems
11. Smith, R. G., and Nolan, B. (n.d.). Manual evaluation of synthesised sign language avatars. Technological University Dublin. https://core.ac.uk/download/482671536.pdf
12. International Organization for Standardization and International Electrotechnical Commission. (2023). ISO/IEC 42001:2023 Information technology, Artificial intelligence, Management system. https://www.iso.org/standard/81230.html
13. International Organization for Standardization and International Electrotechnical Commission. (2025). ISO/IEC 42006:2025 Requirements for bodies providing audit and certification of artificial intelligence management systems. https://www.iso.org/standard/42006
14. Kazakh-Russian signing avatars targeted to native deaf signers [Poster]. (n.d.). sign-lang@LREC. https://www.sign-lang.uni-hamburg.de/lrec/pub/26040_poster.pdf
15. Loyens and Loeff. (2026). EU AI Act: Key amendments adopted under Digital Omnibus package. https://www.loyensloeff.com/insights/news--events/news/eu-ai-act-key-amendments-adopted-under-digital-omnibus-package/
16. Comparative analysis of sign language interpreting agents perception: A study of the deaf. (2024). Proceedings of LREC-COLING 2024, 3603 to 3609. https://aclanthology.org/2024.lrec-main.319.pdf
17. Hickman, L., Hagerty, A., and Powell, R. (2025, September). [Report on British Sign Language AI procurement in UK public services]. Minderoo Centre for Technology and Democracy; British Deaf Association. https://bda.org.uk/wp-content/uploads/2025/10/MCTD-BSL-Report-WEB-02.pdf

18. MultiLingual. (2026). CoSET SAFE AI Task Force proposes a risk framework for automated interpreting at SLxAI Summit. https://multilingual.com/coset-safe-ai-task-force-risk-framework-automated-interpreting-slxai-summit/
19. National Law Review. (2026, July 31). EU Digital Omnibus on AI enters into force. https://natlawreview.com/article/eu-digital-omnibus-ai-enters-force
20. Evaluation of avatar translations into Sign Language of the Netherlands (NGT). (2021). Proceedings of the 1st International Workshop on Automatic Translation for Signed and Spoken Languages (AT4SSL). https://preview.aclanthology.org/moar-dois/2021.mtsummit-at4ssl.9.pdf
21. OECD.AI. (2026). Catalogue of tools and metrics for trustworthy AI: EUD Artificial Intelligence and Sign Language. https://oecd.ai/en/catalogue/tools/framework-for-informed-consent/tool-use-cases/eud-artificial-intelligence-and-sign-language
22. Regulation (EU) 2026/1744 amending Regulation (EU) 2024/1689 (Digital Omnibus on AI), in force July 27, 2026. Summary: https://www.euaiact.com/digital-omnibus-ai
23. Oncins, E. (2026, July). Artificial intelligence driven sign language interpreting avatars [Presentation]. ITU Kaleidoscope 2026. https://www.itu.int/en/ITU-T/academia/kaleidoscope/2026/Documents/Presentations/P1-presentation.pdf
24. Interpreting SAFE-AI Task Force. (2023, September 18). Interpreting SAFE-AI Task Force launches with a mission to guide ethical use of AI in language interpretation. Newswire. https://www.newswire.com/news/interpreting-safe-ai-task-force-launches-with-a-mission-to-guide-22131026
25. Interpreting SAFE-AI Task Force. (2024). First-of-its-kind ethical AI guidance for interpreting released by SAFE-AI Task Force. Newswire. https://www.newswire.com/news/first-of-its-kind-ethical-ai-guidance-for-interpreting-released-by-22376003
26. SAFE AI Task Force and Coalition for Sign Language Equity in Technology. (2025, August 27). SAFE AI Task Force and CoSET release AI Interpreting So-

lutions Evaluation Toolkit, Part A. Newswire. https://www.newswire.com/news/safe-ai-task-force-and-coset-release-ai-interpreting-solutions-22630867

27. Signapse. (2026). AI governance framework and Translation Quality Assurance Policy (policy dated February 9, 2026). https://signapse.ai/ai-governance-framework
28. techUK. (2026). One year on: The Trusted Third-Party AI Assurance Roadmap [Event, September 28, 2026]. https://www.techuk.org/what-we-deliver/events/one-year-on-the-trusted-third-party-ai-assurance-roadmap.html
29. World Federation of the Deaf and World Association of Sign Language Interpreters. (2018, March 14; updated April 14, 2018). WFD and WASLI statement on use of signing avatars. https://wfdeaf.org/wp-content/uploads/WFD-and-WASLI-Statement-on-Avatar-FINAL-14032018-Updated-14042018.pdf

Tips

NCG welcomes confidential tips about sign language AI deployments, procurement decisions and vendor practices at tips@novaracg.com. For readers in the United Kingdom: writing to NCG is not, by itself, a protected disclosure under the Public Interest Disclosure Act 1998. If you are considering whistleblowing about your employer, independent advice is available from Protect at protect-advice.org.uk.

> **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](https://buy.stripe.com/8x2bIV5qH7PP580aeu1wY00). NCG is a for-profit company, not a charity. Contributions are not tax-deductible and buy no service or influence.

Jak cytować

Grizzle, H. M. (2026, November 14). Who Checks the Signing? AI Assurance Is Growing Up Around Sign Language AI. Novara Consulting Group. https://www.novaracg.com/pl/2026/11/14/who-checks-the-signing-ai-assurance-is-growing-up-around-sign-language-ai/