NOVARA
CONSULTING GROUP

This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2026/10/25/the-record-of-what-was-said/.

The Record of What Was Said

Heather Grizzle · October 25, 2026

Can an avatar replay replace English-literacy checks in sign-to-text AI? Only if it replays what the system captured, not what it concluded.

Contenuti

1. The word that came back yellow
2. Where verification lives in SL2T today
3. Translation is not verification
4. Two loops that look the same on screen
5. What the trace keeps and what it drops
6. A replay is still a visual channel
7. Three objects, three claims
8. What the evidence would need to show
9. What buyers should require
10. Rispondere alla domanda
11. How this piece was checked
12. Riferimenti

IN BREVE

- Sign-to-text AI fails fluently. Google DeepMind's own launch material shows "prey" coming back as "grey" and the area code 207 coming back as 2007, and in my own use of SL2T the sign WHY has come back as YELLOW. Each error is a real English word, and nothing on the screen announces it.
- SL2T's safeguard is the signer reading the English and fixing it on a keyboard. The developer and its advisory committee say plainly that this presumes functional English literacy, and that without it "translation errors may pass unnoticed."
- An avatar that signs the system's English back to the signer is back-translation, not verification. If the system heard YELLOW, the avatar signs YELLOW. Translation research has long found that round-trip checks miss real errors and raise false ones.
- An avatar driven by the landmarks the system captured before translation is a different object: a capture trace. It can show the signer what the machine kept, and it can locate where meaning changed. In SL2T, where the video is discarded at once, that trace is the only surviving copy of what the signer did.
- The trace is still thin. SL2T conditions on 130 two-dimensional key points out of the 553 landmarks MediaPipe Holistic can detect, with no tongue, no environment and no depth ordering. A replay can verify the trace without verifying the signing.
- Buyers should require a capture replay where they can get one, forbid calling a generated back-translation "verification," and keep the hierarchy intact: the machine's English is a proposal, the replay is an inspection path, and the signed source, or a qualified interpreter, remains the authority.

The word that came back yellow

In my own use of SL2T, some signs come back less reliably than others, and one of them is WHY. On repeated use, WHY has at times come back on the screen as YELLOW. This is not a benchmark result and I will not present it as one. It is one signer's repeated observation of one system, and it proves nothing about SL2T's overall accu-

racy. What it shows is the shape of a particular kind of failure, and that shape is the subject of this piece.

YELLOW is a valid English word. It is spelled correctly, it sits comfortably in a sentence, and nothing about it signals that a mistake has occurred. A hearing reader on the other end of a text message would see a slightly odd sentence and move on. A Deaf signer who reads English easily would catch it and fix it. A Deaf signer who does not read English easily, and there are many, including signers whose first and strongest language is ASL and who were educated in systems that never gave them full access to written English, has no visible reason to stop it. The sentence goes out under their name.

Google DeepMind's own launch material documents the same pattern in its own words. Its 12 August 2026 announcement of SL2T lists the errors that remain: rare signs, rapid fingerspelling, where "prey" became "grey," passive constructions, classifier depictions, where the word "claws" was dropped, and tense without context, where "kicked off" became "start" [1]. The joint impact report DeepMind wrote with its AI Sign Language Advisory Committee (AISLAC), whose co-authors include representatives of the National Association of the Deaf, RIT/NTID, DPAN and the World Federation of the Deaf, records a model that "consistently" distorted particular number sequences, "systematically transcribing the Maine area code 207 as 2007" [2]. Every one of those errors produces something that reads as English. That is what makes them dangerous.

Fluent errors on the public record of SL2T

Fonte	What was signed or meant	What the English said	Why it passes unnoticed
DeepMind launch post	fingerspelled "prey"	grey	a real word in a plausible sentence
DeepMind launch post	a classifier depicting claws	the claws dropped	the sentence still reads
DeepMind launch post	kicked off	start	tense lost but grammatical
AISLAC report (P0-2)	area code 207	2007	a plausible year
Author's own use	WHY	YELLOW	a real word with no warning

Source: Google DeepMind (2026); AISLAC Joint Impact Report for SL2T 1.0 (2026); author's own use, reported as an observation and not a measured result

This piece continues an exchange. In August I published a governance reading of SL2T, "The Coordinates of Trust" [4]. Jeff Panasuik answered it with "The Second Language Condition," an architectural reading arguing that SL2T removed gloss but kept a written-English output target that decides what counts as success and what becomes the record [3]. He closed by asking which safeguard gives an institution something it cannot reinterpret, negotiate away or quietly ignore. I answered that question on 8 October: the safeguards an institution cannot quietly ignore are the ones that belong to the Deaf person, meaning a right to a qualified interpreter and the signed source as the record [5]. I am not repeating that answer here. This piece takes up a different question raised in the public discussion of Jeff's article, one that answer left open: if English literacy is the precondition, can the precondition itself be changed by letting the signer check the machine's interpretation in sign?

Where verification lives in SL2T today

It is worth being exact about how SL2T is designed to be checked, because the design is more careful than most of the field and its limits are stated by the developer itself.

SL2T runs as sign-to-text dictation in Gboard, anywhere a user would type, and as signed replies in Live Transcribe, on Pixel 11 at launch, from ASL into English [1]. The impact report names the design principle directly: "the signer remains directly in the loop as a verifier, and the model functions as an assistive draft-generation tool rather than an autonomous interpreter" [2]. In practice that principle has two parts. The generated English is shown to the user, who can review it and make corrections "using the standard touch keyboard" before sending it, and in Live Transcribe nothing is shown or voiced to the other person until the signer releases it.

Then the report adds a note that deserves to be quoted in full, because it is the most honest sentence in the release: "This human-in-the-loop safeguard presumes that the user possesses functional English literacy to verify the generated text and sufficient device proficiency to operate on-screen editing controls. Where these conditions are not met, translation errors may pass unnoticed" [2].

HOW AN SL2T MESSAGE IS CHECKED, ACCORDING TO THE DEVELOPER'S OWN DOCUMENTS

Sign
The signer signs to the phone's camera.

→ Track

Track
MediaPipe Holistic, running on the device, tracks points on the signer's face, body and hands.

→ Send coordinates

Send coordinates

Only the coordinates go to the server, and the original video can be discarded immediately.

→ Translate

Translate
A transformer model generates English directly from the coordinate sequence, without glosses.

→ Read

Read
Does the signer, reading English, judge the text correct?

Yes → Send No → Type a fix

Type a fix
The signer corrects the English on the touch keyboard.

→ Send

Send
The English goes out, or is shown to the other person, under the signer's name.

Read the flow as a sequence of tests the signer has to pass. The verification step is a reading test in English. The correction step is a writing test in English. The release decision is taken on the basis of both. The system translates out of ASL, and every check on that translation happens in the language it translated into. That is the point Jeff made architecturally, and it is the point the report makes procedurally when it says errors "may pass unnoticed." Both are describing the same structure from different sides.

DeepMind and AISLAC also drew the boundary in the right place for this version. The report confines SL2T 1.0 to low-stakes uses such as messaging, search and casual one-on-one exchanges, and it explicitly prohibits clinical consultations, police interroga-

tions and court proceedings, IEP meetings and classroom interpretation, job interviews and disciplinary hearings, and benefits determinations [2]. It states that the model "must not be used by third parties to evade statutory obligations under accessibility laws" and that it does not believe SL2T 1.0 satisfies reasonable accommodation obligations under the ADA or Section 504. That scoping is a governance achievement and I said so in August. The question here is what would have to be true before the scope could responsibly widen, and verification is the first item on that list.

Translation is not verification

The distinction between translation and verification is the most useful thing in Jeff's article. A system can produce English that scores well against a reference translation while failing to preserve what the signer meant. DeepMind reports a zero-shot score of 70 BLEURT on the FLEURS-ASL benchmark, which it describes as significantly higher than any previously reported score [1]. BLEURT is a learned measure of how closely generated text resembles a reference text. It is a reasonable instrument for comparing models, and it cannot see the signer. It measures the distance between two pieces of English. Whether either piece of English is what a particular Deaf person meant at a particular moment is a question the metric was never built to answer, and the sign language translation literature has been saying for years that the field's evaluation habits overstate what its numbers mean [11,12].

Verification asks a different question: did the meaning, the direction, the referents, the force, the uncertainty and the intent present in the signing survive the transformation? If the only evidence offered for survival comes from the same process being evaluated, survival has not been independently established. The system has been asked to certify itself, in the language it was optimized to produce.

That is why the obvious fix is weaker than it looks. If English is the problem, translate the English back into sign and let the signer check it. Translation research has tested this idea for decades under the name round-trip or back-translation, and the results are not encouraging. Harold Somers tested round-trip translation as a quality check on machine translation and found it "a poor predictor of quality," with back-translations scoring well even when the forward translation was bad [9]. Dorothée Behr's study of back-translation in survey research reached the same conclusion from the other direction: it leaves many real problems undetected and raises a substantial number of

false alarms [10]. A second fluent output tends to inherit the first system's assumptions, and where it does not, it introduces its own.

> **A CORRECTION TO OUR OWN REQUIREMENT**
>
> In September NCG published six requirements for sign-to-text systems in "The Words Put in Your Mouth." The third was back-translation to the signer before anything is recorded, so that the Deaf person sees, in their own language, what the system is about to say in their name [6]. The goal was right and the mechanism was underspecified. A back-translation generated from the system's English output shows the signer the system's conclusion, which is useful for catching disagreement but is not independent evidence. This piece narrows that requirement: what the signer should see is a replay of what the system captured, before translation, and the replay should never be described as verification of meaning.

Two loops that look the same on screen

The idea that an avatar could change the verification problem came up in the public discussion of Jeff's article. SignsMeta proposed animating an avatar directly from MediaPipe landmark data rather than asking a generative system to produce a plausible sign, so that the avatar shows the user what the system actually retained. I have not been able to retrieve a permanent link to that comment and I credit it here from my own reading of the thread. The proposal deserves a careful answer, because on a phone screen two very different designs would look almost identical.

In the first design, the system translates the signing into English, or into some internal representation of meaning, and then generates signing from that output. This is back-translation with a face on it. It may be a good interface, and it has one real virtue: a signer who sees the avatar sign YELLOW knows something went wrong without having to read a word of English. But the avatar has not established what the system perceived or where the error entered. If the first transformation turned WHY into YELLOW, the avatar faithfully signs YELLOW. It renders the system's conclusion in a second form, and a generative avatar will render it fluently, because fluency is what generative systems are built to produce.

In the second design, the avatar is animated directly from the landmark data captured before translation. It is not producing a sign that matches the system's conclusion. It

is replaying the movement information the system retained on its way to that conclusion. That is a different evidentiary object, and it supports a different kind of reasoning.

Back-translation avatar	Capture replay
Generated from the system's English output	Driven by the landmarks captured before translation
Shows the system's conclusion	Shows what the system kept on the way to its conclusion
Can repeat the first error faithfully	Can expose the layer where the error entered
Fluent by design, whatever went wrong	Only as clear as the capture was
Tells the signer the output disagrees with them	Tells the signer and an evaluator which part of the pipeline failed
Evidence about the translation's result	Evidence about sensing and representation

The diagnostic value is easiest to see with the error that started this piece. If the replay visibly shows the WHY I intended while the English reads YELLOW, the information survived capture and the error happened later, in the model's decoding or in the language modelling that pulled it toward a familiar English word. If the replay itself shows something that is not WHY, because the handshape collapsed, the movement was lost or the face did not register, the problem is earlier: in sensing, in landmark extraction, or in what the representation can hold.

READING A WHY THAT CAME BACK YELLOW, WITH A CAPTURE REPLAY

Watch the replay

Does the replay, driven by the captured landmarks, show the sign the signer intended?

Yes → Decoding error **No → Capture error**

Decoding error

The movement survived capture. Meaning changed during translation into English.

→ Model fix

Model fix
A translation or language-model problem, owned by the model developer and testable against the retained trace.

Capture error
The trace itself lost or changed the form. Translation never had the right input.

→ Sensing fix

Sensing fix
A sensing, lighting, framing or representation problem, which may be fixable by the signer in the moment or only by redesign.

Those two branches lead to different fixes and different accountable parties, which is exactly what an incident log, a correction right or a procurement remedy needs to know. A back-translation avatar cannot split them. A capture replay can.

There is a further reason the second design matters for SL2T in particular. DeepMind's privacy design sends only coordinates to the server and allows "the original video to be discarded immediately" [1]. That is a genuine privacy protection, and signers' faces and bodies deserve it. But it has a consequence the release does not discuss. Once the video is gone, the landmark sequence is the only surviving copy of what the signer did. The impact report even includes an illustration, its Figure 1, of what the model "sees" for a single frame [2]. A capture replay is that figure in motion. It is not a new artifact the developer would have to invent. It is the evidence the architecture already keeps.

Once the video is di-

What the trace keeps and what it drops

scarded, the landmark trace is the only surviving copy of what the signer did.

A capture replay is only as good as the capture, and here the public record is unusually specific. MediaPipe Holistic, the on-device tracker SL2T uses, is a published Google component. Its current documentation describes 478 face landmarks, 33 pose landmarks and 21 landmarks for each hand, 553 in all, with pose and hand landmarks available in three-dimensional world coordinates [7]. Google Research's original announcement described the same topology as "540+ keypoints," using an earlier 468-point face model [8].

SL2T does not use all of that. The impact report states that the model's input "consists of 2D landmark coordinates for 130 key points covering the face, body, and hands" [2]. The report does not say which 130 points are used or how they divide among the face, the body and the hands. What it does say is direct about the consequences: the face model "does not track landmarks on the tongue, which is a critical articulator in ASL"; the representation "captures no information about contents of the signer's environment, which can be used for referencing real-world entities"; and "the lack of depth ordering means the distinction between contact and hovering can be difficult to perceive" [2].

WHAT THE CAPTURE TRACE HOLDS

553
landmarks MediaPipe Holistic can detect on face, body and hands

130

key points SL2T's model conditions on

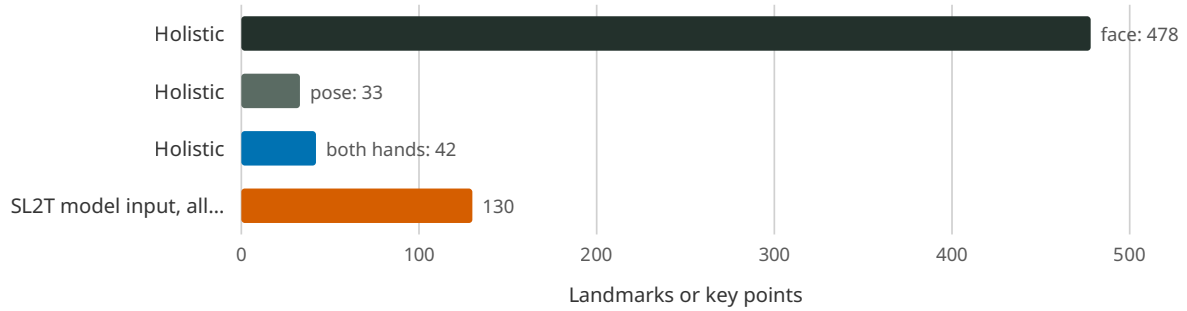
2D

coordinates in SL2T's input, so no depth ordering

0

landmarks on the tongue, a critical ASL articulator

Source: MediaPipe Holistic Landmarker documentation (2026); AISLAC Joint Impact Report for SL2T 1.0 (2026)



The first three bars are what the current Holistic Landmarker can output. The last is SL2T's input. The impact report does not say how its 130 key points divide among face, body and hands.
Source: MediaPipe Holistic Landmarker documentation (2026); AISLAC Joint Impact Report for SL2T 1.0 (2026)

View data

Landmarks or key points	Value
Holistic	face: 478
Holistic	pose: 33
Holistic	both hands: 42
SL2T model input, all regions	130

None of this is a criticism of the decision to use landmarks. The report argues, reasonably, that the representation is efficient, privacy-protective and, with adequate training, rich enough for the uses it targets. The point is narrower and matters for verification. A replay reconstructed from 130 two-dimensional points can only show what those points hold. If a meaning was carried by the tongue, by a pointing gesture to a real object in the room, or by whether one hand touched the other or stopped just short of it, the replay cannot show it, and it will not show its own absence. It will reproduce the omission with perfect fidelity.

The non-manual problem sharpens this. The report's own list of limitations says the model "sometimes misses critical Non-Manual Markers, such as facial expressions, eyebrow positioning, and head nods or shakes, which convey negation, questioning, and core grammatical affect," and its P1-2 issue records that "lexemes conveyed pure-

ly with NMMs often fail to translate entirely" [2]. A replay built from sparse facial points may therefore look like a clean sentence that has lost its question or its negation, which is the most dangerous kind of clean.

What a capture replay of SL2T's trace could and could not show

Feature of the signing	In SL2T's trace	What a replay could show
Handshape and hand movement	yes through hand key points	the form the system captured including a collapsed or lost handshape
Facial grammar such as brows and head nods	partly and inconsistently according to the report	a face that may look flatter than the signing was
Tongue	no tracking at all	nothing and no warning that something is missing
Contact versus hovering	ambiguous without depth ordering	contact and near-contact that look the same
Pointing to real objects	no environmental information	a point aimed at empty space

Source: AISLAC Joint Impact Report for SL2T 1.0 (2026); MediaPipe Holistic Landmarker documentation (2026)

So the stronger design still does not prove that meaning survived. It establishes what one layer of the system retained. That is a real advance over a polished back-translation, because it separates perception from translation and puts the machine's evidence in front of the person whose language it is. It is not the same as verifying the signing, and nobody should be allowed to describe it that way.

A replay is still a visual channel

There is a second limit, and it is personal. I am DeafBlind, and I cannot see above eye level. A replay avatar placed in the upper part of a phone screen, rendered at low contrast or played at native speed, is a replay I may not be able to use. Other DeafBlind signers have different fields, different contrast needs, different preferred modes, and some use tactile signing, for whom a visual replay is not a verification route at all.

This is not an argument against replay. It is an argument against making replay the new universal precondition. A safeguard that moves the burden from reading English to watching an avatar has improved access for many signers and left others exactly where they were. The verification route has to be adjustable, with speed, scale, contrast, framing, repetition and close-up control, and there has to be another route when the replay is not accessible: a human interpreter, a tactile interpreter, or a Deaf reviewer the signer trusts.

The law already points the same way. Under the ADA's Title II effective communication rule, a public entity must give "primary consideration to the requests of individuals with disabilities" when choosing an auxiliary aid or service [15], and a qualified interpreter is one "able to interpret effectively, accurately, and impartially, both receptively and expressively" [16]. A replay feature does not satisfy either provision and should not be marketed as though it did. DeepMind and AISLAC say the same of SL2T itself.

Deaf organizations have also been clear about avatars for a long time. The 2018 joint statement of the World Federation of the Deaf and the World Association of Sign Language Interpreters cautions against signing avatars as a replacement for human signers, particularly where information is live, complex or of significant importance, and accepts them only for pre-recorded static information where "deaf people have been involved in advising on the appropriateness of the signed sentences" [14]. A verification replay is a different use from the one the statement had in view: it shows a signer their own captured signing rather than delivering someone else's message. But the principle carries over directly. Whether a replay is legible and trustworthy is a judgement fluent Deaf signers have to make, and the research on signing animation supports that caution, since facial expression measurably changes how well ASL animation is understood [13].

The evaluation gaps in the impact report matter here too. SL2T was "neither trained nor formally evaluated on signers under 18," formal evaluation across signers with motor disabilities "remain[s] limited," and benchmark disaggregation across sociolects such as Black ASL "remains an ongoing research priority" [2]. Whether a capture replay helps those signers detect errors is unknown, because whether the capture works well for them in the first place has not been shown.

Three objects, three claims

The design question becomes clearer once we stop treating every output as the same kind of record. A sign-to-text system that takes verification seriously handles three different objects, and each supports a different claim.

The signed source is the original communicative event, ideally kept as video when keeping it is lawful, proportionate, consented to and right for the context. It carries the richest evidence of what the signer produced, although interpreting it may still require a qualified human. The capture trace is the machine-readable representation retained from that event: landmark, motion and timing data. An avatar driven directly from the trace lets a signer or an evaluator inspect what the system kept and what it discarded. It is an audit artifact, not the original. The generated text is the model's conclusion about meaning. It may be accurate, useful and editable. Until it is verified, it is a draft, and fluency does not promote it into a record.

Three objects in a sign-to-text system

Object	What it can establish	What it cannot establish	Status in SL2T as publicly documented
Signed source (video)	what the signer actually produced	meaning without competent interpretation	discarded immediately after landmark extraction
Capture trace (landmarks)	what the system retained before translation	features the representation never held such as the tongue or depth	130 2D key points sent to the server; retention period not stated in the public documents
Generated text (English)	the system's proposal of meaning	that the proposal is what the signer meant	shown as a draft and corrected by keyboard

Source: NCG analysis; SL2T status from Google DeepMind (2026) and the AISLAC Joint Impact Report (2026)

Those distinctions change evaluation as much as they change interface design. A serious assessment compares all three objects rather than scoring only the last one, and Deaf-led testing can then ask separate questions. Did the sensing layer preserve the linguistically relevant signal? Did translation preserve the proposition and the intent? Could the signer detect a change and contest it? Was the verification route one this signer could actually use? Did the verification status stay attached to the text after it left the device?

The hierarchy among the three is the safeguard. The machine's English is a proposal. The replay is an inspection path. The signed source, or a qualified interpreter's rendering of it, is the authority. In "The Record Belongs to the Signer" I argued that the signed source should be the record of what a Deaf person said [5]. SL2T's privacy design shows why that argument needs one more step. Where the source is lawfully discarded, the trace is the strongest evidence left, and a system that keeps neither the source nor a trace it can connect to the text has no basis for treating its English as a record of anything.

What the evidence would need to show

The capture replay is a strong design proposition, and so far that is all it is. Preservation may matter more than plausibility. That claim now needs evidence, and the evidence has a definite shape.

1. **Measure correspondence**

   Compare replays against source video for the same signing, rated by fluent Deaf signers who did not produce it, across lighting, framing, signing speed and occlusion. The question is how closely the replay matches what was signed, feature by feature.

2. **Map what survives**

   Test which grammatical features survive the 130-point, two-dimensional representation: handshape, location, movement, orientation, brows, mouth, head movement, contact and use of space. Publish the features that do not survive alongside those that do.

3. **Test error localization**

   Seed known errors and measure whether signers using the replay can tell a capture failure from a translation failure, and how often they are right.

4. **Test correction across literacy**

   Compare how many errors signers catch and correct with the replay, with English review alone, and with both, across a range of English literacy, so that the result shows whether the replay actually lowers the literacy precondition.

5. **Test access configurations**

   Include DeafBlind signers and test speed, scale, contrast, framing and repetition controls, with a tactile or human route recorded as the outcome where the replay does not work.

6 **Test beyond the demo**

Run the tests with real signers across regions, registers and sociolects, including Black ASL, and in conversations rather than isolated sentences, and report results disaggregated rather than averaged.

Those questions do not diminish the idea. They are how an idea becomes assurance, and Deaf-led acceptance testing of exactly this kind is what the SLAT Index Evaluation Standard asks for before any deployment where an error can follow a person into a record [17].

What buyers should require

SL2T 1.0 is out of scope for the settings where these requirements matter most, and its developer says so. But sign-to-text will not stay in coffee shops, and other vendors will not all scope themselves as carefully. These are the terms a public body, an employer or a health system should put in a contract for any system that turns signing into text.

MUST Machine-generated English MUST be labelled as draft output until it has been verified, and the label MUST travel with the text when it is copied, exported, transmitted or passed through an API.

MUST NOT An adverse decision about a person MUST NOT rest on unverified machine output alone.

MUST The signer MUST be able to correct, contest or withdraw the output, and MUST be able to call on a qualified interpreter route, with primary consideration given to the signer's choice.

SHOULD Where the system works from landmarks, the vendor SHOULD provide a replay driven directly from the captured pre-translation data, so that the signer can inspect what the system retained.

MUST NOT A vendor MUST NOT describe an avatar generated from the system's own output as verification of meaning.

MUST The vendor MUST disclose which features its capture represents and which it does not, including face coverage, depth, tongue and environment, so that a buyer knows what a replay can and cannot show.

MUST Any replay MUST be adjustable for speed, scale, contrast, framing and repetition, and MUST NOT be the only verification route offered.

MUST The system MUST keep provenance connecting each piece of text to the trace it came from, and to the signed source where retaining the source is lawful and consented to, for as long as the text can be relied upon.

MUST Where the system cannot connect its text to an authoritative source or trace, the text MUST NOT be filed as a record of what the person said.

MUST Go-live MUST depend on Deaf-led acceptance testing with disclosed scenarios and error categories, and an independent oversight function MUST hold the authority to suspend deployment.

These are not accessories around the model. They decide whether an architectural limitation stays a limitation or becomes an institutional harm, and they are the part of the system a buyer actually controls.

Rispondere alla domanda

Would the option to replay the system's captured signing through an avatar solve the problem of verification in written English? It could materially reduce it, under four conditions: the avatar is driven from the pre-translation capture trace rather than generated from the output; the trace is rich enough to hold the linguistically relevant information for the use in question; the signer can compare the replay with what they meant; and another verification route remains available when the replay is incomplete or inaccessible.

It would not, by itself, establish that meaning survived. It would establish what one layer of the system retained. In SL2T as documented, that layer is 130 two-dimensional points with no tongue, no environment and no depth ordering, and the replay would faithfully show every one of those gaps as if they were not there. That is still a major advance, because it gives the signer the machine's evidence in sign instead of the machine's conclusion in English, and because it tells an evaluator where meaning changed. It is an inspection path, not a verdict.

The final safeguard is therefore not the avatar. It is the rule governing the relationship among the three objects. The machine's English is a proposal. The replay is an inspection path. The signed source remains the authority, and when an institution cannot preserve that hierarchy, it cannot honestly claim to know what was said.

How this piece was checked

METHOD AND DISCLOSURE

Novara Consulting Group is independent of Google, of SignaVision Solutions, of SignsMeta and of every company named here, and holds no financial interest in any sign language AI vendor. The WHY and YELLOW example is my own repeated observation in ordinary use; it is reported as an observation and is not a performance measurement. Statements about SL2T's design, input representation, limitations and scope were checked against Google DeepMind's launch post and the AISLAC Joint Impact Report on 25 October 2026, and landmark counts against Google's MediaPipe documentation on the same day. The SignsMeta proposal is credited from my reading of the public LinkedIn discussion of Jeff Panasuik's article; I could not retrieve a permanent link to the comment. A full SLAT review of SL2T applying the published criteria to the complete public record remains in preparation.

Riferimenti

1. 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/
2. Google DeepMind Sign Language Team, Forbes, S., Kiego, N., Behm, G., Riggio, J., Koraneu, Y., Moore, J., and Thibodeau, R. (2026, August 12). AISLAC Joint Impact Report for SL2T 1.0. Google DeepMind. https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/putting-sign-language-ai-into-users-hands/aislac-joint-impact-report-for-sl2t-1-0.pdf
3. Panasuik, J. (2026, August 26). The Second Language Condition: An Architectural Reading of SL2T. LinkedIn. https://www.linkedin.com/pulse/second-language-condition-architectural-reading-sl2t-panasuik-m-sc--5ukic/
4. Grizzle, H. M. (2026, August 23). The Coordinates of Trust. Novara Consulting Group. https://www.novaracg.com/2026/08/23/the-coordinates-of-trust/
5. Grizzle, H. M. (2026, October 8). The Record Belongs to the Signer: A Reply to Jeff Panasuik. Novara Consulting Group. https://www.novaracg.com/2026/10/08/the-record-belongs-to-the-signer-a-reply-to-jeff-panasuik/

6. Grizzle, H. M. (2026, September 11). The Words Put in Your Mouth: Sign-to-Text AI and Fluent Error. Novara Consulting Group. https://www.novara-cg.com/2026/09/11/the-words-put-in-your-mouth-sign-to-text-ai-and-fluent-error/
7. Google AI Edge. (2026). Holistic landmarker task guide. MediaPipe Solutions, last updated October 1, 2026. https://developers.google.com/edge/mediapipe/solutions/vision/holistic_landmarker
8. Grishchenko, I., and Bazarevsky, V. (2020, December 10). MediaPipe Holistic: Simultaneous face, hand and pose prediction, on device. Google Research. https://research.google/blog/mediapipe-holistic-simultaneous-face-hand-and-pose-prediction-on-device/
9. Somers, H. (2005). Round-trip translation: What is it good for? In Proceedings of the Australasian Language Technology Workshop 2005 (pp. 127-133). https://aclanthology.org/U05-1019/
10. Behr, D. (2017). Assessing the use of back translation: The shortcomings of back translation as a quality testing method. International Journal of Social Research Methodology, 20(6), 573-584. https://doi.org/10.1080/13645579.2016.1252188
11. Müller, M., Jiang, Z., Moryossef, A., Rios, A., and Ebling, S. (2023). Considerations for meaningful sign language machine translation based on glosses. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 682-693). https://doi.org/10.18653/v1/2023.acl-short.60
12. Bragg, D., Koller, O., Bellard, M., Berke, L., Boudreault, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., and Morris, M. R. (2019). Sign language recognition, generation, and translation: An interdisciplinary perspective. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (pp. 16-31). https://doi.org/10.1145/3308561.3353774
13. Huenerfauth, M., Lu, P., and Rosenberg, A. (2011). Evaluating importance of facial expression in American Sign Language and Pidgin Signed English anima-

tions. In Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility. https://doi.org/10.1145/2049536.2049556

14. World Federation of the Deaf and World Association of Sign Language Interpreters. (2018, March 14; updated April 14). 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
15. U.S. Department of Justice. 28 CFR 35.160, Communications. https://www.ecfr.gov/current/title-28/chapter-I/part-35/subpart-E/section-35.160
16. U.S. Department of Justice. 28 CFR 35.104, Definitions ("qualified interpreter"). https://www.ecfr.gov/current/title-28/chapter-I/part-35/subpart-A/section-35.104
17. Grizzle, H. M. (2026). SLAT Index Evaluation Standard (Version 2.0). Novara Press. https://doi.org/10.5281/zenodo.21537271

Come citare

Grizzle, H. M. (2026, October 25). The Record of What Was Said. Novara Consulting Group. https://www.novaracg.com/it/2026/10/25/the-record-of-what-was-said/