



Ambient Voice Technology Buyers' Guide

Understanding AVT assurance in 2026

By Dr Dom Pimenta, CEO at TORTUS AI



If you're buying Ambient Voice Technology (AVT) for your practice, you may now be carrying more responsibility for assuring it than you realise.

The regulatory landscape for AVT has changed. Under current MHRA guidance, technology intended solely to transcribe and summarise clinical conversations, draft correspondence or suggest codes for clinician review may sit outside medical-device regulation. AVT that supports diagnosis, treatment or automated clinical action remains regulated.

That distinction has an important consequence for NHS buyers.

A product can legitimately be available for use in the NHS without its performance and safety having been independently assessed through medical-device regulation.

That does not make the product unsafe. But it does change the questions you need to ask.

If the regulator hasn't checked, who has?

NHS England is clear that responsibility does not disappear simply because a product falls outside medical-device regulation.

The organisation deploying AVT is expected to carry out its own due diligence and maintain appropriate clinical safety, information governance and oversight. It should ensure staff understand the limitations of the technology and review its outputs. It must monitor risks and benefits after implementation and be able to respond when the technology changes.

The clinician remains responsible for reviewing and approving AI-generated information before it is relied upon for patient care. And the supplier still has responsibilities for the safety, security and performance of the product it provides.

Assurance, therefore, is shared. But the responsibility for deciding whether the evidence is good enough to deploy the technology sits with the NHS organisation buying it.

That makes understanding how a supplier has been assured increasingly important.

A listing, certification or supplier claim can tell you something. Independent assessment tells you something else. They should not be treated as interchangeable.

Before you buy, understand three things:

What has been checked?

Performance, clinical safety, cyber security, data handling or simply eligibility against a particular set of criteria?

Who checked it?

The supplier itself, the NHS organisation buying it, an independent assessor or a medical-device Approved Body?

What happens after it changes?

AI does not stand still. Who checks that performance has not deteriorated when the model or functionality is updated?

Start with, "what is this product actually designed to do?"

A tool designed only to produce a draft clinical note sits in a different regulatory position from one intended to support diagnosis, treatment or automated clinical action. But practices should think beyond the label.

Ask what happens downstream from the output. Is the note added to the clinical record? Does it suggest codes? Could later clinical decisions rely on what the AI captured or omitted?

That is why understanding the intended purpose matters more than simply looking for a badge.

Self-declaration, independent assessment and medical-device status

These terms are often presented side by side, but they are not equivalent.

Class I medical-device status has historically involved manufacturer self-declaration rather than the same level of external pre-market assessment required for higher-risk device classes. TORTUS AI argue that this is an important distinction when buyers are assessing how much independent scrutiny sits behind an AVT product.

Class IIa medical devices require external assessment by an Approved Body as part of the UKCA process, alongside a quality-management system and ongoing post-market obligations. TORTUS became the first AVT to achieve UKCA Class IIa certification in June 2026.

That does **not** mean every non-medical-device AVT is unsafe. However, it does mean the buyer should ask: **If an external regulator or auditor has not validated these claims, who has?**

Look at what is being measured

"98% accurate" sounds reassuring until you ask what 98% actually means.

Does the measure capture:

- Omissions?
- Invented information?
- Drug names?
- Diagnoses?
- Clinical codes?
- Structure and formatting?
- Clinically significant errors?

A supplier should be able to explain not just its headline performance figure, but why it chose that measure and why it matters clinically. Do not stop at the claim. Ask to see the evidence.

Ask how many consultations were assessed, where testing took place, who performed the evaluation and whether the results reflect real-world clinical use.

And perhaps most importantly: Ask to see how the system fails, not just how it succeeds. A strong supplier should understand the limitations of its technology just as well as its strengths.

Assurance has to continue after deployment

The most important change in AI procurement is that assurance can no longer be treated as a one-off event. That matters because AI systems change.

A model may be replaced. Prompts can be adjusted. New functionality can be added. Performance can improve in one area and deteriorate in another.

So ask your supplier what happens when the product you originally assessed changes.

Three risks worth understanding

Hallucinations:

AI can produce information that sounds plausible but was not actually present in the consultation.

The current concern around AI scribes is not hypothetical. Recent reporting following NHS safety investigations has heightened concerns about errors in AI-generated clinical information, including drug names and diagnoses. It's important to ask how hallucinations are defined, measured, detected and reviewed.

Model drift:

A system that performed well when tested may not behave identically after model updates or other changes. Ask what benchmarks are rerun after updates and what level of change would trigger additional clinical safety assessment.

Reliance bias:

As clinicians become accustomed to AI-generated notes, there is a risk that review becomes less rigorous. NHS England says organisations should ensure users understand that AVT outputs may contain inaccuracies and must be reviewed before being relied upon for patient care.

Training therefore cannot simply cover how to turn the product on. It should include how to use it safely over time.

The AVT Buyer's Guide

Six questions every supplier should be able to answer

1. How do you measure performance?

Ask exactly what the supplier measures and how. Do not accept "accuracy" without understanding what counts as an error.

Follow up: Which mistakes are included in that figure, and which are not?

2. Why did you pick those measures?

A supplier chooses the metrics it reports. Ask why those measures reflect real clinical risk and performance rather than simply presenting the product favourably.

Follow up: What clinically important problem could your headline metric miss?

3. What evidence do you have that it performs at that level?

Ask for the evidence behind the number. Was the system tested in real clinical settings? How large was the dataset? Was there independent evaluation?

Follow up: Who checked the evidence besides you?

4. How do you monitor that it maintains that level of performance in the real world?

Pre-deployment testing is only the beginning. Ask how live performance is monitored, how often it is reviewed and how customers are informed if performance deteriorates.

Follow up: What would trigger you to pause or roll back the system?

5. How do you ensure system changes don't degrade performance?

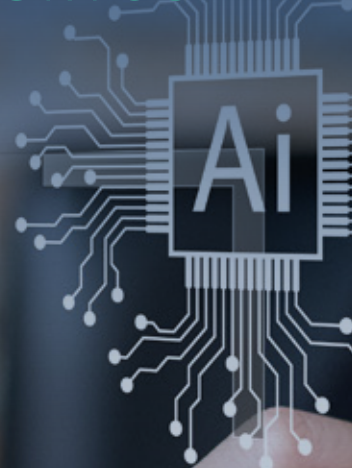
Models and features will change. Ask how updates are validated, which tests are repeated and whether significant changes receive independent or regulatory scrutiny.

Follow up: Will we know when the underlying model changes?

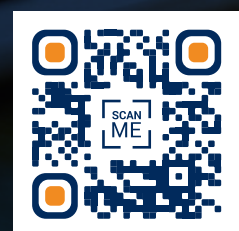
6. Who is responsible when it goes wrong?

Do not leave this until an incident occurs. Understand what responsibility sits with the supplier, what sits with the clinician and what sits with your organisation.

Follow up: Walk us through what happens after we report a clinically significant error.



Choosing AVT should begin with the right questions. See the evidence behind Surgery Intellect, powered by TORTUS, and how it supports confident adoption in general practice.



Find out more

