

AI QUALITY ASSURANCE 2026

AI Call Scoring: QA on Every Call | VOCPhone

Your best team leader spent four hours last week listening to eleven calls, then had awkward conversations about three-week-old ones nobody remembered. AI can review all nine hundred.

Manual call quality assurance was never good, it was just the only option available. Reviewing eleven calls out of nine hundred is statistically hollow, arrives weeks too late to coach, reads as a verdict, and burns the most leveraged labour in the room. AI changes coverage from a sample to **everything**, which turns quality from an opinion into a data set. Be precise about capability, though: **reliable** on whether things were said, on mechanics like dead air and talk-over, and on classifying every call; **partly reliable** on sentiment; **not reliable** on whether the answer was right for that customer. Three failure modes: **transcription error that isn't randomly distributed**, **measuring what's easy instead of what matters**, and, the big one. **a team that experiences it as surveillance**.

The short version

Time spent	On what	What came out of it
2 hours	Listening to eleven recorded calls, chosen because they were the right length and easy to find.	Eleven scorecards. Coverage: about 1.2% of the week's conversations.
1 hour	Filling in the scorecards and writing feedback notes.	Documentation that will be read once.
1 hour	Six short conversations about calls from two to four weeks ago.	Two people couldn't remember the call. One disputed the sample. Three nodded.

■ Also worth knowing

Manual review	Keyword spotting	AI evaluation	
Coverage	~1.2% of calls	100%, phrases only	✓ 100% of calls
Latency	Days to weeks	Immediate	✓ Immediate
Understands meaning	✓ Fully	✗ Not at all	Substantially
Consistent call to call	✗ Varies by assessor and mood	✓ Perfectly	✓ Highly
Cost per call reviewed	High, senior labour	Negligible	✓ Low
Trusted by the team	Depends on the leader	✗ Openly gamed	Depends entirely on rollout

Two things that cut the other way

First, **recording and monitoring calls has its own rules**. In Australia there's no single national rule, call recording is governed by state and territory listening-devices and surveillance-devices legislation, which varies, on top of your Privacy Act 1988 and Australian Privacy Principles obligations. The safe posture almost everyone adopts is simply to tell people: a clear notice at the start of the call plus a documented internal policy. Get that right *before* you build scoring on top of it.

■ Questions the full guide answers

- ☐ What is actually wrong with sampling a few calls per agent?
- ☐ What can AI judge reliably, and what can it not?
- ☐ How do we stop the team seeing this as surveillance?
- ☐ Is the compliance case strong enough to justify it on its own?
- ☐ Can we use AI scores in performance reviews or pay decisions?
- ☐ If we run an AI Phone Agent, does it need reviewing too?
- ☐ What do we need before we can start?

This is the summary. The guide has the detail.

Every point above is worked through in full on our site, with the Australian context, the numbers and the exceptions. If you would rather talk it through, call 1300 663 222 or visit vocphone.com/about/contact-us.