

AI TRIAL PROTOCOL · 2026

The Ten Calls That Pick Your AI Phone Provider

Four quotes, four demos that all went well, and no way to tell them apart. Ten calls you make yourself, scored out of twenty, sort a four way tie in one afternoon.

A demo is a rehearsed conversation between two cooperating parties. A business call is nothing of the sort. Your callers interrupt, mumble a suburb from a moving car, change their mind halfway, and ring at ten past nine at night because something has gone wrong. **Never accept a demonstration in place of a trial number you can ring yourself, at times you choose, for a week.**

■ In the demo, and then on Tuesday

In the demo	On Tuesday
A quiet room and a good connection.	A ute on the highway, or a car park with two bars.
A caller who waits for the prompt.	Someone talking over the greeting because they are in a hurry.
Clear, deliberate speech.	A mumbled surname, and a number said as "double oh, four two, triple three".
Questions it was built to answer.	Something you have never published anywhere.
Eleven on a Tuesday morning.	Ten past nine on a Saturday night.
It ends at the summary.	The summary has to become a record somebody acts on.

■ The ten calls, in this order. About ninety minutes per provider

- 1 **The interrupt.** Talk over it mid sentence. Does it stop and follow you, or finish its script?
- 2 **The local proper noun.** A suburb you serve, a customer surname, one of your product names. Then read the transcript.
- 3 **The mumbled mobile.** A callback number said fast and naturally, ideally from a car.
- 4 **The wrong department.** Ask for one thing, then change halfway. Does what you said travel with you?
- 5 **The unanswerable question.** A service you do not offer. It should defer, not invent.
- 6 **The emergency.** Describe something urgent in plain words, without using the word emergency.
- 7 **The 9pm call.** After hours, and again on a Saturday.
- 8 **The changed mind.** Book, change the day, change it back, then ask what you have booked.
- 9 **The integration check.** Look in your own CRM, not their dashboard.
- 10 **The repeat caller.** Ring back an hour later as the same person. Does it remember?

■ Score each 0, 1 or 2, immediately, before discussing it with anyone

Call	0, fails	2, genuinely good
Interrupt	Talks over you	Stops and follows the new point
Proper nouns	Suburb and surname both wrong	Right, and custom vocabulary can be loaded
Mumbled number	Wrong digits, no flag	Right, and read back for confirmation
Re-route	Start again from nothing	Re-routes with everything you said
Unanswerable	Invents an answer	Says it does not know, offers a person
Emergency	Keeps collecting details	Recognises urgency and connects immediately
After hours	Undefined behaviour	Hours, holidays and a genuine after hours path
Changed mind	Books the wrong thing	Handles the reversal in conversation
Integration	Nothing in your CRM	Right contact, right fields, within a minute
Repeat caller	No recognition	Recognises and continues the thread

Two patterns show up almost every time

First, the spread is much wider than the price spread. Providers within ten per cent of each other on cost routinely land eight or nine points apart out of twenty. Second, the calls that separate them are never the ones in the sales deck. Calls 1, 5 and 9 do most of the discriminating, and none of the three appear on any feature comparison table published anywhere. Run the ten calls before the discount arrives, while everything is still hypothetical.

■ Three questions to settle in writing before signing

- **Where is the audio processed, and where does it rest?** Different things, sometimes different countries. Ask whether your audio trains anybody's models.
- **How long is it kept, and who can retrieve it?** Per record type retention, deletion that deletes, role based access, and a tested bulk export.
- **Who can see the call when something is wrong?** Ask how many companies sit between your fault and the person who can fix it.

The two questions to end every trial with

"What do the first ninety days look like, week by week?" A specific answer names who does what and when, with a test day, a go live day and a thirty day review. Vagueness there predicts exactly the implementation you are imagining, and more good products fail on implementation than on capability. Then: "Who do I ring at 4pm on a Thursday, and what can they see?" The second half matters more. Reaching a person is table stakes. Reaching a person who can look at your actual call, your routing and the AI behaviour in one place is the difference between a fix and a ticket.