Free guide · Updated July 2026

The AI Receptionist Buyer’s Guide

Everything to check before you put an AI on your phone line — useful whether you pick Sarah or anyone else.

AI receptionists went from novelty to a real purchasing category fast, and the market is now crowded with products that sound identical on their landing pages: answers 24/7, sounds human, books appointments. Underneath the identical claims are very different products — different voice quality, different booking depth, radically different pricing models, and different amounts of work left on your desk. This guide is written to help a service-business owner evaluate the category properly, whichever vendor you end up choosing.

We build one of these products — Sarah247 — so we are not neutral bystanders, and we will use Sarah as a concrete example where an example helps. But the checklist, demo questions, and red flags below are the ones we would hand a friend evaluating any vendor, including us. The goal is that you finish this guide able to sit through any demo and know exactly what to poke at.

What an AI receptionist actually is

An AI receptionist is software that answers your business phone line with a natural, conversational voice, holds a real dialogue with the caller, and takes actions: asking your qualifying questions, answering questions from your business’s knowledge, booking appointments into your calendar or field-service software, texting callers, and escalating to humans by your rules. It attaches to your existing phone number — typically via call forwarding — so nothing changes on the caller's side except that the phone always gets answered.

The load-bearing word is actions. The generation of tools before this — voicemail, phone trees, message-taking services — captured information and pushed the work back to you. The defining property of a real AI receptionist is that a call can go from ring to booked job with no human in the loop, and the humans review the recording afterward rather than doing the work in the moment.

What it is not

Clearing up the category's baggage is worth a minute. This is not the robotic phone tree that made everyone hate automated phones — there are no menus and nothing to press; callers just talk. It is not a chatbot bolted onto a phone line; voice conversation with interruptions, accents, and background noise is a distinct and harder problem, which is why voice quality varies so much between vendors.

It is also not a replacement for your judgment. An AI receptionist executes the playbook you give it — your prices, your triage rules, your escalation paths. Vendors implying the AI will figure your business out on its own are describing a product that does not exist. The realistic promise is different and better: your best phone process, executed identically on every call, at any hour, without sick days.

When you actually need one

The clearest signal is structural: calls arrive when nobody can answer them. You work in the field and the phone rings during jobs; your office covers forty hours of a 168-hour week; leads come from Yelp or Thumbtack where the fastest responder wins. If any of that describes you, missed calls are not an occasional accident — they are a standing feature of how your business is shaped, and every one is a job that likely went elsewhere.

The honest counter-cases: if your phone rarely rings, fix marketing before answering infrastructure. If a full-time person already answers every call to your satisfaction and after-hours volume is genuinely negligible, your problem may already be solved — though it is worth pulling the call log before trusting that impression. And if most calls require deep human judgment rather than intake and scheduling, look at hybrid setups where AI handles overflow and after-hours only.

Voice quality: the first thing to judge

Voice quality is not cosmetic. Callers who realize they are talking to a clumsy robot hang up, and every hang-up is a lead lost — so the naturalness of the voice directly gates everything downstream. Judge latency above all: how long after you stop talking does the response start? Awkward gaps kill conversations faster than any accent. Then judge interruption handling — real callers talk over the agent constantly — and composure with background noise, since your callers phone from job sites, cars, and kitchens.

Never judge this from the vendor’s demo reel. Call the line yourself and behave like a difficult customer: interrupt mid-sentence, change your mind about the appointment time, mumble, ask the same thing twice. A voice agent that stays coherent through that will survive your real call traffic. One that only shines in scripted demos will not.

Booking depth: messages versus scheduled jobs

This is the widest quality gap in the category, hidden behind the identical claim books appointments. At the shallow end, the AI takes down a preferred time and sends you a message — meaning your staff still does the actual scheduling, and you have purchased an expensive voicemail. At the deep end, the AI reads real availability from your systems and writes a real appointment: correct job type, correct duration, correct service-area rules, confirmation text sent, no human touch.

Probe for the difference with operational questions. Does it see true availability or guess? Can it handle your arrival windows and job durations? What happens when the caller wants a slot that is taken? Can it reschedule an existing appointment, or only create new ones? Any hedging here means message-taking with extra steps — which is fine only if that is what you meant to buy.

Integrations: where the appointment has to land

An AI receptionist that books into a calendar nobody at your company looks at has not booked anything. The system of record matters: if you dispatch out of Housecall Pro, jobs must land in Housecall Pro with the fields your dispatcher needs; if your pipeline lives in GoHighLevel or your week lives in Google Calendar, same logic. Ask specifically about your stack, and ask to see a booking appear in it during the demo — not on a slide.

Look past the appointment, too. Where do call recordings, transcripts, and captured lead details go? Can the system fire your missed-call text-back and follow-up sequences? The pattern to avoid is the standalone dashboard that becomes one more inbox nobody checks. The receptionist should feed the tools you already run your business from, not compete with them.

Human handoff and escalation

No AI should handle every call, and the vendors worth trusting say so unprompted. What you are evaluating is the escape hatch: can the system transfer a live call to a human when your rules say so? Can you define those rules — angry customer, emergency by your written definition, caller asks for the owner, question outside trained knowledge? What happens when the human does not pick up: is there a fallback, or does the caller dead-end?

Equally revealing is behavior at the edge of knowledge. Ask the demo agent something it cannot know and watch what happens. The right behavior is honest and graceful — let me take your number and have someone get back to you on that.

The wrong behavior is inventing an answer to seem capable. An AI that guesses about your prices or policies is not an employee; it is a liability with a pleasant voice.

Recordings, transcripts, and oversight

Every call your AI receptionist takes should be recorded and transcribed, and you should be able to review them without friction. This is not surveillance — it is the management layer. Recordings are how you verify quality in week one, catch knowledge gaps, settle he-said-she-said disputes with customers, and mine intake details the summary missed. A vendor that cannot show you full recordings and transcripts is asking you to take call quality on faith.

Check the practical details: how fast after a call are recordings available, are transcripts searchable, is there a digest of after-hours activity waiting in the morning? And ask about consent handling for call recording, which varies by state — a vendor operating seriously in the US will have a ready answer rather than a pause.

Pricing models: per-minute, per-seat, flat

Three models dominate. Per-minute pricing charges for talk time — Sarah247, for example, pairs a base plan from $350 a month with usage from $350 a month plus 15 cents per minute — so cost tracks usage closely: slow season costs little, busy season costs more and is paying for itself in booked jobs. Per-seat or per-agent pricing imports a human-software convention that fits this category badly; you are buying call coverage, not licenses, and a seat that idles all night costs the same as one answering constantly.

Flat monthly plans buy predictability, and predictability has real value — but read the fine print, because most flat plans are minute bundles wearing a costume: overage rates above the cap, wasted spend below it. The comparison that cuts through all three models is cost per outcome. As illustrative math: at from $350 a month plus 15 cents per minute, a four-minute call that books a 300-dollar job cost 60 cents. Run the same arithmetic on every quote you receive, including setup fees amortized over your first year.

Questions to ask on a demo

Bring a written list and make the salesperson demonstrate rather than describe. The essentials: Can I call the agent right now and try to break it? Show me a booking landing in my actual scheduling software. What exactly does onboarding involve, who does the work, and how long until live? How do I change a price or policy after launch, and how fast does the change take effect? What are your escalation options when the AI is out of its depth?

Then the commercial layer: What is the all-in cost — setup, monthly, per-minute, overages? What is the contract term, and what does leaving cost? Who owns my number, my recordings, and my data if I cancel? Can I run a limited trial on my real overflow or after-hours line first? A good vendor answers all of this crisply. Evasion on any of these questions is itself an answer.

Red flags

Long contracts on an unproven service — the product either books jobs within weeks or it does not, and a vendor demanding a year up front is pricing in your inability to leave. Large setup fees for what is mostly configuration work. Per-seat pricing, which signals a product philosophy mismatched to call coverage. Message-only output dressed in booking language — if every call still ends with your staff doing the scheduling, you bought a very expensive notification system.

Softer flags worth weighing: no way to hear the agent live before signing, demo environments suspiciously unlike your trade, vague answers about which scheduling systems are actually integrated versus on the roadmap, and no clear story for human escalation. None of these alone is disqualifying; two or three together tell you the category is being sold to you faster than it is being built.

Implementation and measuring ROI

A competent implementation is measured in days, not months: an onboarding session or two to capture services, prices, intake questions, calendars, and escalation rules; a build-out of that knowledge; test calls with you playing difficult customers; then a soft launch — typically forwarding after-hours or overflow calls first, expanding as trust builds. Expect a tuning period of a couple of weeks where you review recordings and file corrections. A vendor with a longer runway than that should be able to say precisely why.

Then measure like an owner, not a fan. Before launch, note your baseline: calls received, calls missed, after-hours outcomes, booked jobs per week. After launch, track what the AI answered, what it booked, what it escalated, and read a sample of recordings monthly. The core comparison is simple: minutes billed versus jobs on the calendar that used to ring out. If that ledger is not obviously positive within the first month or two, the product is not doing its job — and with no contract in the way, you can act on that conclusion. That, ultimately, is the buying criterion this whole guide reduces to: pick the vendor confident enough to be judged monthly.