

DEPLOYMENT PLAN · 90 DAYS

Your First AI Agent

Week one is not building. Week one is writing down the one job, the things it may never do, where a call goes when it stops, and the sentence that says whether it worked.

■ Week 1: four documents, no software

Document	Test that it is good enough
The job, in one paragraph, including what is out of scope	It fits in a paragraph. A page means the job is too big and should be split.
The never list, flat prohibitions with no conditions	Every item is enforceable by withholding a capability, not by asking nicely.
The handover: trigger, destination, what carries across	A named destination for every hour of the week, including 2am.
The success sentence: one number, one date	Somebody who disagreed would know they disagreed. "Improve experience" fails.

Pick the job on three variables, and get the same answer on all three

High volume, a few dozen a week, because repetition is the only way you learn anything. Low variance, because agents demo brilliantly and behave inconsistently where the work has many shapes. Low consequence, because your process for catching mistakes does not exist yet. The complicated job eating a senior person's week scores badly on all three. It will be a good candidate in a year.

■ The never list. Seven that cover most businesses

- **No price** that is not on the published list. A price on a recorded call is a representation and it belongs to the business.
- **No commitment to a timeframe.** It may state a booked appointment. It may not predict one.
- **No account change without verification,** and the failure path is not an easier second question.
- **No other customer or job,** enforced by constraining the lookup tool, not the wording.
- **No claiming to be a person.** Disclosure up front produces fewer complaints.
- **No handling a caller in distress.** Trigger words, straight to a human.
- **No third attempt.** Two tries to understand, then hand over.

Enforce it in the tools, not the wording

An instruction that says "do not issue refunds" is a preference. **Not giving the agent a refund tool is a control.** Where the consequence is real, take the capability away rather than asking for restraint. Save instruction-level rules for what you cannot express structurally.

■ The rest of the ninety days. Exit tests, not dates

- 1 **Week 2, connect it.** Read the customer record, calendar and jobs; write notes, tasks and bookings back. Only a quarter of contact centres using AI have actually integrated it.
- 2 **Weeks 3 to 4, shadow.** Live calls, nothing reaches a customer, no record changes. Exit when a transcript rarely surprises you.
- 3 **Weeks 5 to 7, after hours only.** Calls that would have hit a mailbox. Tests whether escalation works at 2am, which is a roster question.
- 4 **Weeks 8 to 11, supervised overflow.** Daytime, past a queue threshold, half an hour of review a day inside somebody's workload.
- 5 **Weeks 12 to 13, narrow front line.** Answers first on the defined job only. Autonomy widens one decision at a time.

■ Five numbers

- **Containment.** 40 to 70% is healthy. Near 100% means handover is broken.
- **Resolution.** Sample and score fifty calls a month.
- **Repeat contact in 72 hours.** Flat or falling.
- **Handover quality.** Above 90%.
- **Cost per resolved contact,** review time included.

■ Six questions for the vendor

- Which tools can it call, and can I add mine?
- Show me a transcript you did not script.
- What does it do when unsure?
- Where is my data processed, and what is kept?
- Containment and repeat contact, both.
- What happens to what I build if I leave?

Report containment and repeat contact as one line, never two

A caller told something unhelpful at 7pm who rings back at 9am counts as a success on most vendor dashboards, and has cost you two contacts instead of one. **Rising containment alongside rising repeat contact means the agent is deferring work, not resolving it.** This is the most common way a deployment looks good for two quarters while costing more than it saves.

Tell your team in week three, not week thirteen

Show them the shadow transcripts. It is much harder to be suspicious of something you have already read a hundred pages of. Give them one click to flag a bad call, and **act on the flags visibly**: two or three flags producing a change in the same week beats any amount of explanation. And be straight about what it means for the work, because the version that causes trouble is the one where nobody says anything.