



Evaluating AI voice agents for outbound calling

What to test, what to ask, and where AI outbound quietly breaks.

WHY THIS GUIDE

AI voice agents demo beautifully and fail quietly

The gap between a slick demo and a production rollout is compliance, routing, latency, and cost — the unglamorous parts. Behind every stalled outbound program is the same hidden overhead: the swivel-chair tax. This is the checklist we wish every buyer used.

The 8 KPIs every outbound program must track

- 1. Connection rate — % of attempts that reach a live person. Gauges dialing efficiency + timing.
- 2. Right-party-contact rate — % of answered calls reaching the actual decision-maker.
- 3. Conversion rate — % of calls that hit the goal (booking, payment, qualification).
- 4. First-call close — resolved in one interaction, no escalation.
- 5. Cost per acquisition — total cost per successful outcome. The AI-vs-manual proof.
- 6. Average handling time — minutes per interaction. Lower = more capacity, if quality holds.
- 7. Engagement rate — % of connects that become real conversations.
- 8. Call quality & outcome accuracy — compliance, intent, sentiment, resolution.

IVA vs. generative AI voice agents

The one-line test: IVAs handle menus. Generative agents handle outcomes.

Legacy IVA / IVR	Finn (generative)
Rigid menu trees ("Press 1 for billing")	Conversational, context-aware ("What can I help with?")
Poor containment, caller frustration	Higher containment and first-call resolution
Slow to deploy, complex to maintain	Pilots live in under 30 days, fast iteration
Can't handle nuance at scale	Books, pays, troubleshoots in real time

The ROI case

Model it with: $(\text{Total Benefits} - \text{Total Costs}) \div \text{Total Costs} \times 100$. Independent studies back the payback:

- Forrester: 331% ROI over three years, payback in under six months.
- McKinsey: 30–45% lower service costs from generative AI in customer service.
- IBM: ~40% call-center cost reduction, 30% higher engagement.
- Baseline cost per call runs \$5–8 fully loaded; containment starts 40–60% and climbs past 80%.

The 10-call test

Before you sign, run ten real calls: one angry, one silent, one wrong-number, one that asks something off-script, one in a second language. A vendor that survives those survives production.

How Finn measures up

- Sub-500ms fused voice pipeline with natural barge-in; 46 languages, hot-swappable mid-call.
- DNC, calling-hours, and consent controls built in; every call recorded, transcribed, scored.
- Rules-based warm transfer with full context; writes back to your CRM. SOC 2 / HIPAA / GDPR / ISO 27001.
- Usage-based pricing — you pay for talk, not for ring tones. ROI in 60–90 days.

Run the 10-call test with Finn — www.hirefinn.ai/bookademo

Or start free at hirefinn.ai · hello@hirefinn.ai