

Financial

SEND NO

NEVER FULLER

DEMO DID +1 (650) 476-2005 ONLY

WEB RENDER · LEADS REDACTED

AT A GLANCE

Pricing tiers, per-minute cost stack, unit economics, margins, usage guardrails.

Updated 2026-09-20 07:12 PT · full detail in the sections below — jump straight to one:

1. Pricing (decided)

2. Per-minute cost stack (default config)

3. Per-call and fixed costs

4. Margin at \$199/mo

5. Usage guardrails

6. Startup costs to customer #1

7. Two-line summary

Download PDF ↓

DETAILS

2026-09-19 (PST) · Sage@miniPC · LIVING Derived from: `cost-model.md` (Phase 0 step 0.5; Retell + Twilio public pricing fetched

2026-09-19). Estimates, not quotes — re-check before any pricing change.

Standing locks: no purchases without Barry · demo DID +1 (650) 476-2005 only.

1. Pricing (decided)

TIER	PRICE	NOTE
National	\$199/mo	The pitch price: "one extra completed job pays the month"
Bay Area	\$249/mo	Density/willingness premium
Test city	\$99/mo	Pilot only — margin still fine (see §4)
Floor	\$59/mo	Never discount below this, anywhere

No free trial. Voice only is priced; text assistant / web app / website are "contact us" until Barry sets the menu (SELLING-SYSTEM gap G7).

2. Per-minute cost stack (default config)

COMPONENT	CHOICE	\$/MIN
Retell voice infra	required	0.055

COMPONENT	CHOICE	\$/MIN
TTS	Retell platform voices (not ElevenLabs at \$0.040)	0.015
LLM	small tier (\$0.008; premium option \$0.025)	0.008-0.025
Telephony	US via Retell/Twilio	0.015
Total		\$0.093-0.110

Planning number: \$0.12/min (headroom for add-ons at +\$0.005/min each and price drift). Speech-to-speech realtime models (\$0.07-0.38/min extra) are NOT in the plan — the flow-based agent is fine for intake.

3. Per-call and fixed costs

- Average after-hours intake call: **3 min** (range 2-4) → **≈ \$0.36/call** at planning rate (\$0.28 typical).
- **10-min max call duration** on every agent — bounds ramblers and the two-AI-talk failure mode.
- Fixed per customer per month: Retell DID **\$2.00** + Twilio-direct toll-free SMS (**\$2.15/mo** + ~\$0.02 per 2-segment owner summary) **≈ \$4.15/mo**. Never Retell SMS at \$20/mo — Twilio direct is ~10x cheaper at our volumes.
- Concurrency: first 20 concurrent calls free — a non-issue until ~20+ customers.

4. Margin at \$199/mo

CALLS/MO	MINUTES	VARIABLE (@\$0.12/MIN + SMS)	FIXED	TOTAL COST	MARGIN	MARGIN %
30	90	\$11.40	\$4.15	≈ \$16	\$183	92%
100	300	\$38.00	\$4.15	≈ \$42	\$157	79%
300	900	\$114.00	\$4.15	≈ \$118	\$81	41%

- Break-even at \$199: ~1,620 answered min/mo (~540 three-minute calls) — comfortable.
- At the \$99 test-city price: 100 calls/mo ≈ \$42 cost → **\$57 margin (58%)** — fine for pilot.
- Customer value anchor: 33 booked jobs/mo × \$500 avg = **\$16,500 captured** vs \$199 cost (≈ 82× before the first invoice).

5. Usage guardrails

1. **Alert at 300 answered minutes/customer/month** (matches ToS fair-use line) → fleet alert to Barry, then a pricing conversation — never a silent overage bill.
2. **Hard cap: 10-min max call duration** on every agent (set at provisioning).
3. Monthly cost-per-customer line in the daily report once customer #1 exists.

6. Startup costs to customer #1

- Production DID: ~\$2-5/mo (Barry authorizes purchase — money rule).
- Stripe + Twilio accounts: \$0 to open (Barry identity/bank; Batch #1).
- Toll-free verification: fast path, low cost; A2P 10DLC: registration fees, needs entity/EIN — started in Batch #1, runs in background.
- Everything else (agents, hosting, monitoring) rides existing infrastructure $\approx$ \$0 incremental.

7. Two-line summary

100 calls/mo costs $\approx$ **\$42**; margin **\$157 (79%)** at \$199. A heavy shop at 300 calls still clears **\$81 (41%)** — pricing is safe with the 10-min cap + 300-min alert.

Web-version note: specific lead business and owner names are redacted here to “Lead A-K (metro, vertical)” labels. Full names live in the private repo registry (`agents/leads/leads-data.json`) and the live **leads pipeline**.

Generated 2026-09-20 07:12 PT from `receptionist/docs/financial.md` by `bin/publish-docs.py`. Republish: `python3 ~/apex/bin/publish-docs.py` then `rsync docs/` per `COS-DEPLOY.md`. Never Fuller.