

Production readiness

SEND NO

NEVER FULLER

DEMO DID +1 (650) 476-2005 ONLY

WEB RENDER · LEADS REDACTED

AT A GLANCE

Frozen 2026-09-19 audit: what works, blockers B1-B5, secondary gaps, execution plan.

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

1. What ALREADY works end-to-end today (don't undersell)

2. Critical blockers (ranked)

3. Secondary gaps (needed soon, not blocking customer #1)

4. Execution plan — today's state → first paying customer

5. Compact version (phone-readable)

Download PDF ↓

DETAILS

2026-09-19 (PST) · Sage@miniPC · read-only audit (no code, no SEND flags, no Trackmint touched)

Question from Barry: **"What do we need to do to roll out in production? What are we missing?"** Scope: docs + code under `~/apex/receptionist/`, `~/apex/website/projects/ai-assistance/`, CoS/leads progress, vault decisions. Standing locks respected throughout: Never Fuller · demo DID +1 (650) 476-2005 only · AUTHORIZE SEND NO.

1. What ALREADY works end-to-end today (don't undersell)

- **Voice agent answers a real phone number.** Retell AI staging agent on the demo DID +1 (650) 476-2005, refreshed daily ("Call 17th" beat confirmed LIVE 2026-09-19). Caller gets greeted, qualified, name + callback captured, next-open slot offered.
- **Web-call prototype** (`receptionist/projects/software-template/retell-flow/`): Node 20 server mints Retell web-call tokens, flow compiled from JSON (`flow/northline.json`), 11/11 tests passing, 0 npm vulns. Template-per-vertical architecture already designed (copy template → client folder).
- **Demo hub is a real sales asset.**
<https://apex.socialtokens.site/projects/ai-assistance/> with 7+ vertical demos (trades, restaurant, laundry, auto, mechanics, dental, law), buyer proof path Shop → Book → Card → Stubs, mobile-first, iPhone frames.
- **Owner broadcast panel** (`demo/broadcast/`): builds + previews subscriber messages; SEND hard-disabled by design (button

disabled, nothing leaves the browser). Safe demo.

- **Pricing is DECIDED:** \$199/mo (Bay \$249 · test-city \$99), no free trial, pitch = "one extra completed job pays the month." (CLIENT-ACQUISITION.md, 2026-09-15.)
- **Lead pipeline machinery works:** Leads → Legal → Marketing → QA → CoS hold → Barry send. **4 Track A leads legally cleared with drafts written and HELD:** DFW (Lead A (DFW plumbing), Lead B (DFW plumbing)) + Houston (Lead C (Houston-metro plumbing), Lead D (Houston-metro plumbing)). Legal reviews are genuinely good (CAN-SPAM/TCPA/DTPA analysis per shop).
- **Ops muscle:** VPS deploy pipeline, daily beat cadence, PR-based cloud agents, fleet monitoring infra (though not yet pointed at the call path — see gap O1).

Honest framing: this is a demo-grade single-tenant product with a production-grade sales pipeline sitting behind a closed gate. `progress.md` self-assesses pctReady 58 — fair.

2. Critical blockers (ranked)

B1 — No production call path for a real customer (BIGGEST)

- **Missing:** Everything is one staging agent + one shared demo DID answering as a fake shop. There is no way to give a real plumber their own number: no per-customer DID purchase, no per-customer Retell agent provisioning, no call-forwarding instructions.
- **Why it blocks:** The product literally cannot serve customer #1.
- **Effort:** M (the template/copy-per-client architecture already exists; needs a provisioning script + one bought DID + a forwarding one-

pager for the shop).

- **Who:** Sage builds provisioning + runbook autonomously; **Barry authorizes the DID purchase** (money = Barry per CLAUDE.md ping rule; ~\$2-5/mo per number via Retell/Twilio).

B2 — Booking promise isn't real: no SMS, no persistence

- **Missing:** The pitch is "we answer, book next open, **text you the window**." Today the SMS to the owner is a stub, bookings live only in the browser session, no calendar/CRM write, no backend store.
- **Why it blocks:** First customer would pay \$199 and discover the core deliverable (owner gets a text with the booked window) never arrives.
- **Effort:** M-L. Minimal viable: Retell post-call webhook → small VPS service → store booking + send owner SMS. **Warning: A2P 10DLC registration for business SMS takes days to weeks and needs an entity/EIN — start it first, or use a toll-free verified number (faster) for the pilot.**
- **Who:** Sage builds the webhook/store/notify service; **Barry signs up the SMS provider account (Twilio) and supplies entity details for registration.**

B3 — No live billing

- **Missing:** Stripe is a "checkout sketch" on the hub; no way to actually charge \$199/mo.
- **Why it blocks:** No payment = no paying customer, by definition.
- **Effort:** S — a Stripe Payment Link / subscription for \$199/mo is under a day once the account exists. Full self-serve checkout is later polish.
- **Who: Barry creates the Stripe account** (identity + bank + entity details — only he can); Sage wires the payment link,

webhook, and receipt flow.

B4 — Production-call compliance pack missing

- **Missing, for the calls themselves (not outreach):**
- **No AI disclosure** in the greeting — `northline.json` greet node never says the caller is talking to an AI/virtual assistant. Several states now require this; it's also the honest default and costs one sentence.
- **No recording disclosure.** Texas + California are one-party consent, but if Retell call recording is on, the greeting should disclose it ("this call may be recorded") — esp. since our first leads are TX and our numbers are CA (two-party state) DIDs.
- **No privacy policy / ToS** for the shop customer or their callers; no documented retention policy for transcripts/recordings.
- **CAN-SPAM postal address still a placeholder** in all held drafts (`[POSTAL ADDRESS – Barry to insert]`) — flagged as QA blocker #1; blocks any send.
- **Business entity / insurance:** not documented anywhere in the repo or vault.
- **Why it blocks:** Real customers + recorded real-caller PII without disclosures = legal exposure from day one; postal address blocks the outreach that finds customer #1.
- **Effort:** S for greeting lines + policy pages + retention doc (Sage drafts all of it).
- **Who:** Sage drafts everything; **Barry supplies: postal/mailing address, entity name (or decision to operate as sole prop for the pilot), and 2 minutes of approval.**

B5 — The gate itself: SEND is NO and only Barry can flip it

- **Missing:** 4 cleared, drafted, QA-able Track A emails are sitting on disk. Zero sent, ever. First customer cannot appear until outreach (or some channel) goes live.
- **Why it blocks:** No top-of-funnel = no customer, regardless of product readiness.
- **Effort:** 5 minutes of Barry, after B4's postal address lands in the footers.
- **Who: Barry decision + action** (authorize send, per-shop by name, per standing rule).

3. Secondary gaps (needed soon, not blocking customer #1)

#	GAP	WHY IT MATTERS	EFFORT	WHO
01	No monitoring on the call path — nothing in <code>~/apex/monitoring/</code> watches the demo DID or a future customer line; no alert if Retell is down or the agent errors mid-call	"Fails as silence on a real client's phone" (our own CLAUDE.md)	S — daily/hourly scripted test call + Retell API health check into existing fleet-health alerting	Sage autonomous
02	No failure fallback — if Retell/LLM dies mid-call, caller gets dead air; no forward-to-voicemail/owner-cell fallback configured	First bad night destroys trust with customer #1	S-M (carrier-side no-answer forwarding config + documented	Sage designs; per-customer config

#	GAP	WHY IT MATTERS	EFFORT	WHO
			in onboarding)	
O3	No onboarding runbook — nothing documents "plumber says yes → live": intake (greeting, hours, escalation number, services, price disclosures), provisioning, forwarding setup (*61*/conditional forward), test call, go-live checklist	Every hour of Barry-time per customer is margin lost	S — Sage writes it	Sage autonomous
O4	No customer-facing transcript/booking visibility — owner gets (future) SMS but no log/dashboard of calls & bookings	Churn risk; also the upsell surface	M — simple per-customer page can ride on existing owner-board demo UX	Sage
O5	Cost/capacity model undocumented — Retell $\approx$ \$0.07-0.13/min + DID + SMS; at \$199/mo a chatty shop could eat margin; no per-account usage cap	Pricing safety	S — one doc + usage alert	Sage autonomous
O6	Multi-tenant backend — per-client folder copies are fine for 1-5 customers; real auth/tenancy/persistence needed to scale	Explicitly listed in progress.md Missing; NOT needed for customer #1	L — defer	Sage, later
O7	Marketing site / SEO / self-serve signup — hub	Inbound channel	M — after first paying	Sage, later

#	GAP	WHY IT MATTERS	EFFORT	WHO
	is demo-grade, no public pricing/checkout page	stays dark	customer	
O8	Calendar/CRM integration (real write to Google Calendar / Jobber / Housecall)	"One shared book" SKU promise; SMS-only window is acceptable v1	M-L — defer, sell SMS-window v1	Sage, later

4. Execution plan — today's state → first paying customer

Principles: batch Barry's actions into **three short sessions**; everything else Sage runs autonomously with proof (burn-in, not assurances). Long-lead items (SMS registration) start on day 1. All money/brand/go-live moments ping Barry per standing rule.

Phase 0 — Compliance + plumbing prep (Days 1-2)

STEP	WHAT	WHO	DURATION	UNBLOCKS
0.1	Add AI + recording disclosure lines to template greeting (all vertical flows), staged on demo agent for verification	Sage	2 h	B4, Phase 3 go-live
0.2	Draft privacy policy, ToS, data-retention policy (transcripts 90 days default, recordings off by default for	Sage	3 h	B4

STEP	WHAT	WHO	DURATION	UNBLOCKS
	pilot) as hub pages, unlinked until approved			
0.3	Write onboarding runbook + customer intake form (greeting text, hours, escalation cell, services, forwarding instructions per carrier)	Sage	3 h	O3, Phase 3
0.4	Stand up demo-DID health monitor (scripted Retell API check + daily test web-call) wired into fleet alerting	Sage	2 h	O1
0.5	Write cost model (per-min Retell + DID + SMS vs \$199) + usage-alert threshold	Sage	1 h	O5
0.6	BARRY BATCH #1 (~30 min, one sitting): (a) supply postal/mailling address for CAN-SPAM footers; (b) confirm entity name (or "sole prop for pilot"); (c) create Stripe account; (d) create Twilio account + start toll-free SMS verification (fast path) and A2P 10DLC (slow path, background); (e) authorize purchase of 1 production DID (~\$2-5/mo)	Barry	30 min	B2, B3, B4, B5

Phase 1 — Make the promise real (Days 2-5, after Barry batch #1)

STEP	WHAT	WHO	DURATION	UNBLOCKS
1.1	Insert postal address into the 4 held drafts; QA pass (incl. Lead D	Sage	30 min	B5 send

STEP	WHAT	WHO	DURATION	UNBLOCKS
	(Houston-metro plumbing) surname check)			
1.2	Stripe: \$199/mo subscription Payment Link + webhook → notify on signup; test-mode proof first	Sage	half day	B3
1.3	Booking persistence: Retell post- call webhook → small VPS service → store booking + owner SMS via Twilio (toll-free) ; burn-in on demo line 48 h	Sage	1-2 days	B2
1.4	Provision "pilot plumber" agent from plumber template on a SECOND staging config: disclosure greeting, real hours, escalation number, end-to-end test (call → booking stored → SMS received)	Sage	1 day	B1
1.5	Report to Barry with proof: recording of test call, screenshot of SMS, Stripe test receipt	Sage	—	Barry confidence for Phase 2

Phase 2 — Open the funnel (Day 5-6)

STEP	WHAT	WHO	DURATION	UNBLOCKS
2.1	BARRY BATCH #2 (~10 min): review 4 drafts, AUTHORIZE SEND per shop by name (Lead C (Houston-metro plumbing), Lead D (Houston- metro plumbing), Lead A (DFW plumbing), Lead B (DFW plumbing))	Barry	10 min	Top of funnel

STEP	WHAT	WHO	DURATION	UNBLOCKS
2.2	Send the 4 Track A emails from Barry's Gmail (one shop per email, per held drafts); log sends; honor opt-outs ≤ 10 days	Sage (with Barry's send auth)	1 h	Replies
2.3	Reply watch: Sage drafts responses within the hour, demo links + demo DID; books a live demo call slot with interested owner	Sage drafts / Barry approves sends	ongoing	Phase 3
2.4	In parallel: run next lead batches through the existing Leads→Legal→Marketing→QA machine (Houston Track B revisit, next metro) so the funnel isn't 4 emails deep	Sage	ongoing	pipeline depth

Phase 3 — First close + go-live (Week 2-4, driven by reply rate)

STEP	WHAT	WHO	DURATION	UNBLOCKS
3.1	BARRY BATCH #3: the human close — one phone/email exchange with the interested owner; take the yes	Barry	~30 min per prospect	Contract
3.2	Customer pays via Stripe link → Sage provisions: buy their DID, clone plumber template with their greeting/hours/escalation, disclosures on, monitor on	Sage	half day	Live service
3.3	Customer sets after-hours conditional forwarding (from our one-pager); joint test call;	Customer + Sage	1 h	Live

STEP	WHAT	WHO	DURATION	UNBLOCKS
	Barry pinged at go-live (brand/live rule)			
3.4	48 h burn-in: monitor every call, Sage reviews transcripts daily, fallback verified; then declare customer #1 LIVE with numbers, not assurances	Sage	2 days	Proof

Phase 4 — After customer #1 (not before)

Multi-tenant backend + auth (O6) · owner dashboard (O4) · calendar/CRM write (O8) · marketing site/SEO/self-serve checkout (O7) · scale outreach beyond hand-held sends.

Barry's total footprint to customer #1: three batches ≈ 70 minutes + one sales conversation. Everything else is Sage, with proof at each gate.

5. Compact version (phone-readable)

Works today: Retell voice agent live on demo DID · 7-vertical demo hub · pricing set (\$199/mo) · 4 legally-cleared outreach drafts HELD · deploy + daily-beat ops.

Blockers: (1) no per-customer number/agent provisioning · (2) booking SMS + storage are stubs — core promise not real · (3) no live Stripe · (4) no AI/recording disclosure, no privacy/ToS, no postal address · (5) SEND gate closed.

Order: Sage prep (disclosures, policies, runbook, monitor) → Barry 30-min batch (address, Stripe, Twilio, DID auth) → Sage wires

SMS+storage+billing, proves it on staging → Barry authorizes 4 sends → replies → Barry closes one → Sage provisions → 48 h burn-in → first paying customer. ETA $\approx$ 2–4 weeks, gated mostly by email reply rate.

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/production-readiness-2026-09-19.md` by `bin/publish-docs.py`. Republish: `python3 ~/apex/bin/publish-docs.py` then `rsync docs/` per `COS-DEPLOY.md`. Never Fuller.