

The WhatsApp Bot Blueprint — Turn Conversations Into Customers on Autopilot

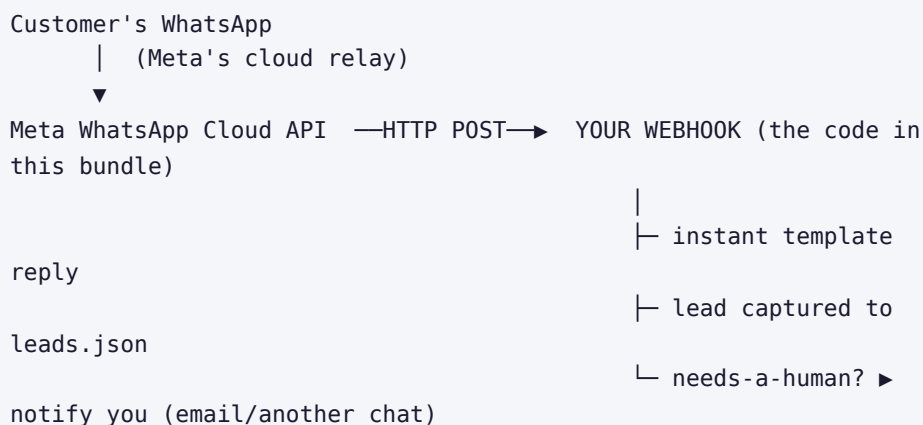
Version 1.0 · 2026 · Includes: Blueprint PDF + Reference Webhook Code +
Go-Live Checklist

1. The 30-Second Version

WhatsApp is where your customers already are. This blueprint walks you from zero to a bot that: answers common questions instantly, captures leads into a list you own, and escalates the conversations a human should handle — running on YOUR server, with YOUR model, no per-message fees.

Deliverables in this bundle: - `blueprint.pdf` — this guide -
`reference/webhook.py` — Meta WhatsApp Cloud API webhook (verify + receive + auto-reply) - `reference/go-live-checklist.md` — the 12-point launch gate -
Single-business commercial license

2. How the Pieces Fit (2 minutes)



Two facts people don't tell you upfront: 1. **You need a small “system prompt” layer in Meta's UI** (pre-approved templates for the first outbound message) — it's free tier, not a purchase. 2. **The webhook must respond < 20 seconds.** Our

pattern: reply instantly with an acknowledgment template, then send the real answer one second later. Customers read it as “typing fast.”

3. Setup Walk-Through (45 minutes)

Step 1 — Meta developer account (10 min). Create at `developers.facebook.com`. Add the “WhatsApp” product. The sandbox panel gives you a test number immediately — build on the sandbox first, migrate to production numbers after your checklist passes.

Step 2 — Point Meta at your server (10 min). Your webhook needs a public HTTPS URL. Fastest local options, in order of permanence: - A server you own with a real domain (best), - A free tunnel for development (e.g. `cloudflared tunnel --url localhost:8080`).

In Meta’s app: WhatsApp → Configuration → Webhooks → set URL + `VERIFY_TOKEN` (any secret string you also put in the code below) → subscribe to the `messages` field.

Step 3 — Drop in the reference code (10 min). `pip3 install flask requests`, save `reference/webhook.py`, run `python3 webhook.py`. Send a WhatsApp message to your test number: you’ll get the auto-reply. Your bot is alive.

Step 4 — Plug in your brain (10 min). The reference code has a `brain()` stub. Replace it with your local model (the AI Tron Starter Kit’s `think()` works as-is) or any API you prefer. The rest of the code — verification, dedupe, lead capture, guardrails — stays.

Step 5 — The business layer (5 min). Edit `BUSINESS_FACTS` near the top: hours, location, returns, anything a customer asks weekly. The bot answers from facts, not imagination.

4. What the Reference Code Actually Does (worth 10 minutes of reading)

1. **GET verification** — Meta’s handshake, per spec, first try.
2. **POST handling** — extracts the customer’s text; ignores status pings and echoes.
3. **Dedupe guard** — Meta retries on timeout; a message-ID cache stops double replies.
4. **Lead capture** — every new number lands in `leads.json`: who, when, what they asked.

5. **3-class router** — *answerable* (facts) / *sales* (mark HOT, notify you) / *human* (complaints, legal, urgent) — the only three outcomes that matter at launch.
6. **Instant ack + staged reply** — pattern from Section 2.

The one customization that earns your money back: the router's sales branch. A lead asking "do you offer X for Y scenario" is a buying signal — the code stamps it HOT and notifies you within seconds. Call within the hour; you'll close more than any ad improvement.

5. The 3 Templates That Do 80% of the Work

Template	When Meta sends it	Example
Welcome	first contact	"Thanks for reaching ACME — how can we help? (hours/repairs/booking)"
Away	outside hours	"We're offline till 9am — leave your question and we'll reply first thing."
Escalation ack	HOT/human class	"Got it — a specialist will call you shortly. Reply here to add details."

Write these three in your voice and your bot already sounds like your best employee.

6. Compliance & Deliverability (read once, forever)

- Opt-in: customers must message you first (reply to inbound = always allowed for 24h via Meta's service window).
- No bulk blasts from the bot — that's an instant rate-limit on Meta.
- Log every conversation. Your future self (and refunds) will thank you.
- Never ask for payment data over chat. Link out to your checkout.

7. Go-Live Checklist (full version in [reference/go-live-checklist.md](#))

☐ sandbox message round-trips ☐ fact answers correct ☐ after-hours template fires ☐ lead lands in leads.json ☐ HOT notification received ☐ dedupe test (re-send) ☐ kill switch known

8. Troubleshooting

- **403 on webhook** → `VERIFY_TOKEN` mismatch or URL not HTTPS.
- **Nothing arrives** → messages field not subscribed (Step 2, last click).
- **Slow replies** → you're calling a large model synchronously; reply with ack first.
- **Duplicate replies** → dedupe cache missed; check message-ID extraction.

Ship the boring version first. Bots earn trust one correct answer at a time.

— *All Is One. Build fast.*