GNOSIS LABS · FREE CHECKLIST

# The Nine Failure Classes of Production Voice Agents

A demo exercises the happy path: a polite caller, a clean request, one tool, a tidy close. Production is a caller who interrupts, a webhook that 502s, a parameter that vanishes between the console and the wire, and a retry that books the slot twice. Nine classes cover the difference. One self-check per class; every one of these has cost somebody a call that mattered, and most of them have cost us one.

## A Transport and tool wiring

The most expensive class, because a simulated test suite reports it green. A passing "Tool Called" assertion means the model chose the tool; it does not mean an HTTP request left the building.

☐ Has one real call, per release, been asserted against the destination database rather than the log? And has anyone probed your endpoints from a public vantage instead of inside your own network?

## B Identity, auth, and secret hygiene

Your tool endpoints are public by construction; the vendor's cloud calls them. Webhooks carry transcripts and PII on the strength of a URL nobody is supposed to know.

☐ Does a bare POST with no credential get refused? Is the webhook signature verified over raw bytes before parsing? Does the 401 avoid teaching the attacker the credential's shape?

## C Idempotency and state races

The caller says yes once. A timeout triggers a retry. Two calendar entries. Or the tool fails, the agent apologizes, the call ends, and no record of the caller survives at all.

☐ Does replaying an identical request body create a second row? Does a failed booking still leave a durable lead behind?

## D Conversation control

Dead air during tool calls, barge-in ignored, emails never read back, latency filler with two owners that stacks five stall phrases before any content.

☐ Is every email and phone number read back, spelled? Who owns the time-buying phrases: the prompt, the platform filler, or both without knowing it?

### E Truthfulness and commitment boundaries

The single most expensive hallucination class is a price the caller repeats back to you in writing. The softer sibling hedges a fixed price into uselessness.

☐ Can your agent quote a figure that is not in the knowledge base? Can it promise anything that binds a human who was not on the call?

### F Adversarial input

"Ignore your previous instructions." The subtler sibling arrives through a tool result: a calendar entry or CRM note carrying instructions, returned into context by the agent's own tools.

☐ Is caller speech treated as content to record, never instruction to follow? Is tool output delimited and labeled as untrusted data?

### G Failure and degradation

"I received a 502 from the booking service" should never reach a caller's ear. A failing tool retried five times in one turn bills five times.

☐ When a tool dies mid-call, what does the caller actually hear? What happens to caller N+1 when the concurrency cap is hit?

### H Data protection and compliance

Recording is on by default in most consoles; several jurisdictions require all-party consent, and in some it is criminal rather than a fine. Retention defaults to forever.

☐ Is recording disclosed in the opening turn, matched to the caller's jurisdiction? Is there a defined retention period and a deletion route that has actually been exercised?

### I Observability and evidence

The post-call webhook fires, the durable write fails, the handler returns 200, and the transcript is gone with no redelivery. Vendors ship new config fields carrying active defaults nobody chose.

☐ Do you monitor the destination table for recency, or the code path for errors? Would you notice if your platform shipped a new behavior into your agent's config last night?

---

**The full audit runs 43 probes across these nine classes: $5,500 fixed, five business days.** Graded report with per-probe evidence, a ranked fix list, and a regression checklist your team keeps. We run our own production receptionist on this discipline: call **+1 (617) 766-0577** and try to break it. · gnosislabs.tech/voice-audit · Cap Dawes, Gnosis Labs