

Voice agent QA checklist (STT+TTS)

Tip: Print this page. Aim to check every box before scaling traffic.

Voice & TTS (trust + first impressions)

- ☐ Test at least 3 voice styles (neutral, friendly, energetic) and compare abandonment in the first 10 seconds.
- ☐ Avoid assuming “more human” is better: validate whether highly expressive / voice-actor-like TTS increases hang-ups.
- ☐ Tune speaking rate, pauses, and prosody for phone audio (clarity over theatrics).
- ☐ Add a fallback voice and a “re-intro” line if early hang-ups spike after a voice change.

STT precision (critical entities)

- ☐ Define a “must-be-correct” entity list (first/last names, addresses, emails, IDs, order numbers).
- ☐ Measure precision on that entity list separately from global WER.
- ☐ Use custom vocabulary / phrase boosts for proprietary terms, brands, and product names.
- ☐ Add confirmation/repair flows for high-risk fields (spell-out, repeat-back, or double-confirm).

Evaluation in real conditions

- ☐ Build a test set that matches production: accents, noise, overlapping speech, and telephone artifacts.
- ☐ Include bilingual and code-switching cases and evaluate on task success, not only benchmark labels.
- ☐ Track business outcomes alongside speech metrics (completion, conversion, churn, CSAT).
- ☐ Review “failure buckets” weekly (misheard entity, latency, barge-in, confusion) and prioritize fixes.

Go-live & monitoring

- ☐ Instrument early funnel metrics (first-utterance hang-up, first-turn latency, first-entity accuracy).
- ☐ Log STT hypotheses and final transcripts for QA, with privacy and compliance controls.
- ☐ Create an escalation path: fallback to a human or offer a call-back when confidence is low.
- ☐ Run a small “canary” rollout and expand traffic only when core metrics are stable.

Inbound vs. outbound call design

- ☐ Create separate configs for inbound and outbound (STT/TTS settings + orchestration rules).
- ☐ Inbound: support broader vocabulary and exploratory intent collection with graceful clarifications.
- ☐ Outbound: optimize for speed and correctness when confirming identity, appointments, and transactions.
- ☐ For outbound, tighten thresholds for proceeding without confirmation on critical entities.

Latency & turn-taking (perceived speed)

- ☐ Set an end-to-end latency budget and instrument each stage (audio → STT → reasoning → TTS).
- ☐ Tune conversational rhythm: too slow feels broken; too fast feels interruptive.
- ☐ Handle interruptions (barge-in): stop TTS quickly, listen, and acknowledge before continuing.
- ☐ Add “repair” behaviors (e.g., rephrase, apologize, ask a clarifying question) when confidence drops

Experimentation as a system

- ☐ Run ongoing A/B tests across voices, speaking styles, turn-taking, and latency thresholds.
- ☐ Define a small set of “north star” metrics (abandonment, completion, resolution, conversion).
- ☐ Ship changes behind flags and roll back quickly when abandonment or error rates spike.
- ☐ Keep a short experimentation backlog (voice, timing, confirmations) and iterate weekly.

Orchestration & UX (treat as a product)

- ☐ Set expectations in the opening line (what the agent can do, and what it will ask to confirm).
- ☐ Design clarifying questions for ambiguity (avoid guessing on names, dates, or amounts).
- ☐ Provide graceful exits (handoff, SMS link, callback) instead of dead-ends.
- ☐ Document your “conversation rules” so product, design, and engineering can iterate together.