

Automated AI Voice Agent QA Pipeline

Continuous Integration for Voice AI Logic & Behavior

 Status: Inactive

This workflow is currently inactive due to third-party API changes, but reflects production-ready automation design and QA logic used to stress-test the "Denver Best Plumber" AI Receptionist.

Workflow Visualization

(Insert n8n workflow screenshot here)

Node Architecture Overview

- Trigger: Scheduled Cron (2x Daily) for regression testing
- Persona Generation (Gemini): Dynamic customer scenarios
- Vapi Caller Trigger: AI Tester calls AI Receptionist
- Webhook Listener: Captures end-of-call payload
- AI Auditor (GPT-4o): Evaluates transcript using rubric
- Logging (Google Sheets): Stores scores and summaries
- Alerting (Gmail): Sends QA report

Call Flow (Adversarial AI Testing Loop)

An AI Tester Agent places calls to the AI Receptionist, creating a continuous adversarial testing loop.

QA Logic & Evaluation Rubric

1. Dynamic Persona Injection

Each test injects a unique persona to prevent overfitting. The tester agent is constrained to complete the booking flow.

2. Strict Judge Rubric

1. Address Readback Rule
2. Emergency Detection & Priority Routing
3. One-Question-Only Data Collection
4. Clear Next Steps & Confirmation Messaging

3. Error Handling & Transcription Logic

The auditor ignores transcription artifacts and filters non-final call events to prevent duplicate or incorrect logging.

Tools & Tech Stack

- n8n: Workflow orchestration & automation logic

- Vapi.ai: Voice AI agents (Tester & Receptionist)
- Google Gemini: Persona generation
- OpenAI GPT-4o: Transcript auditing & scoring
- Google Sheets: Historical QA database
- Gmail: Automated alerts

Key Improvements Achieved

- Improved QA score from 2/10 to 10/10
- Eliminated hallucinated filler responses
- Correct emergency escalation behavior
- Enforced strict single-question data collection