

Cayman Roden

QA Automation Engineer | 12,000+ Tests · pytest · Playwright · CI/CD · LLM Evaluation

310-982-0492 | caymanroden@gmail.com | linkedin.com/in/caymanroden | chunkytortoise.github.io/qa.html | Remote, US-based

Professional Summary

QA Automation Engineer with 12,000+ automated tests shipped across 15 production repositories, all developed TDD-first. Designed test automation frameworks using pytest, Playwright, and Vitest with CI/CD integration via GitHub Actions. Built LLM evaluation pipelines including a 24-fixture golden eval with adversarial test cases, RAGAS scoring, and LLM-as-judge CI gates that block regressions before merge. Full-stack test ownership: test architecture design, framework selection, and coverage strategy from project inception.

Technical Skills

Test Frameworks: pytest | pytest-asyncio | Playwright | Vitest | unittest

Test Types: Unit | Integration | E2E | API Contract | Regression | Adversarial | LLM Evaluation

CI/CD: GitHub Actions | Docker | bandit (SAST) | pip-audit (CVE scan) | mypy strict | ruff | pre-commit

Languages: Python | TypeScript | JavaScript | SQL

Infrastructure Under Test: FastAPI | PostgreSQL | Redis | REST API | WebSocket | GoHighLevel Webhooks

Methodology: TDD-first | Shift-Left Testing | Test Strategy | Test Architecture Design | Quality Assurance | Agile | JIRA | Coverage Enforcement (87-92%)

Test Automation Portfolio

EnterpriseHub - Multi-Agent Orchestration Platform (7,678 tests, 1,100+ CI-verified)

pytest - pytest-asyncio - mypy strict - bandit - pip-audit - GitHub Actions

- Designed test suite covering planner/coder/QA agent mesh: inter-agent handoff verification, per-agent model routing validation, autonomous task delegation correctness
- Cache testing: L1 memory / L2 Redis / L3 PostgreSQL hit/miss/eviction edge cases, TTL behavior, 88% aggregate hit rate validation
- Security testing: Ed25519 webhook signature verification, parameterized SQL injection resistance, rate limiting enforcement, OWASP control validation
- CI pipeline gates: SAST (bandit), dependency CVE scan (pip-audit), type safety (mypy strict on 15 modules), linting (ruff) – all block merge on failure

DocExtract AI - LLM Evaluation Suite (1,183 tests, 87%+ coverage)

pytest - RAGAS - LLM-as-judge - Brier score calibration

- Designed golden evaluation set: 24 fixtures including 8 adversarial cases (contradictory documents, truncated inputs, multi-language, high-noise OCR), 94.6% extraction accuracy baseline
- Built RAGAS evaluation pipeline: context recall (0.35), faithfulness (0.40), answer relevancy (0.25) scored by LLM judge with rubric and evidence extraction
- Regression gate: CI blocks any extraction accuracy drop below baseline before merge
- Confidence calibration: Brier score tracking predicted vs actual extraction success across fixture types

Jorge Real Estate AI - Production SMS Bot Suite (1,824 tests)

pytest - pytest-asyncio - GitHub Actions

- Designed multi-layer test strategy for a greenfield codebase: selected pytest-asyncio for async event-loop correctness; parameterized fixtures cover all routing branches across 3 bots (Lead Intake, Buyer, Seller) without duplication
- Conversation flow coverage: multi-turn state machine verification, bilingual routing logic (EN/ES), qualification scoring thresholds, admin handoff triggers, and cross-bot reassignment edge cases
- Security layer: dual-scheme webhook signature verification (HMAC + RSA), replay prevention, malformed payload rejection
- Integration layer: Redis rate limiting enforcement (60 req/min), atomic session lock verification, state recovery after connection loss

Finance Analytics Portfolio - Cross-Stack Test Architecture (1,611 tests)

pytest - dbt tests - Vitest - API integration tests

- 4 test layers: Python analysis (1,472 pytest), data model (33 dbt schema + custom SQL), frontend component (36 Vitest), API contract (70 integration)
 - dbt schema tests validate mart model data integrity; custom SQL tests catch HMDA source data anomalies
 - API contract tests verify FastAPI response schemas match Zod frontend validators at every boundary
-

Experience

QA Automation Engineer (Contract) | Acuity Real Estate - Remote - January 2026 - March 2026

Sole QA owner for a production AI SMS bot platform serving a real estate client. Established testing standards for a greenfield codebase from day one.

- Set test strategy before any production code merged: framework selection (pytest-asyncio), coverage thresholds, CI gate configuration, and branch protection rules
- Validated client-facing behavior end-to-end: 500+ real leads processed with no routing errors, unhandled exceptions, or client-visible failures across the 3-month deployment
- Shipped 1,824 tests at handoff with complete documentation: test strategy doc, coverage report, and runbook for adding new conversation flows
- Zero production regressions; zero failures across the full deployment lifecycle (January - March 2026)

QA Automation Engineer (Contract) - Remote - June 2022 - December 2025

Designed and maintained test automation frameworks for AI, API, and data analytics applications across 15 production repositories.

- Shipped 11,810+ automated tests using TDD-first methodology: failing test, minimal implementation, refactor, verify. Zero tests written retroactively.
- Built CI/CD pipelines with GitHub Actions including SAST scanning, dependency CVE auditing, type checking, and coverage enforcement gates
- Designed LLM evaluation infrastructure: golden test sets, adversarial fixture generation, RAGAS scoring pipelines, and LLM-as-judge CI gates
- Published mcp-server-toolkit to PyPI with custom MCPTestClient enabling unit testing of MCP servers without live API keys (412 tests, 88% coverage)
- Contributed to LiteLLM (27K+ stars): typed exception mapping for Router fallback, eliminating silent failure in multi-model routing

Service Operations | Hospitality Industry - 2016 - 2022

Quality-focused execution across 50-200 covers per shift; systematic error detection and multi-step service protocol adherence under time pressure.

Certifications

Selected from 21 certifications across Google, IBM, Microsoft, and DeepLearning.AI (1,831 total hours):

- IBM Generative AI Engineering (PyTorch, LangChain, HuggingFace) - 144 hours
- Open Source Software Development, Linux and Git (Linux Foundation) - 60 hours
- Duke University LLMOps Specialization - 48 hours
- IBM RAG and Agentic AI Professional Certificate - 24 hours
- Python for Everybody (University of Michigan) - 60 hours
- Microsoft AI and ML Engineering Professional Certificate - 75 hours