

AYOOLUWA PAUL OBEMBE

AI Systems QA & Evaluation Engineer | LLM/RAG & Agent Reliability | Legal Practitioner

Lagos, Nigeria | +234 805 760 6459 | obembeayooluwa.work@gmail.com

[Portfolio](#) | [GitHub](#) | [LinkedIn](#)

PROFESSIONAL SUMMARY

AI Systems QA & Evaluation Engineer and Legal Practitioner with 2+ years of experience across SaaS, fintech/RegTech and legal AI. Designs CI-ready evaluation harnesses for LLMs, RAG pipelines, tool-calling agents and regulated workflows, covering human-feedback rubrics, hallucination and grounding checks, prompt regression, workflow assertions and root-cause diagnostics. Built TestAlgnite and LexTutor, and designed a comparative benchmark across five AI systems and eight high-stakes legal-reasoning tasks using a nine-dimensional, 45-point rubric.

CORE EXPERTISE

LLM & Agent Evaluation: model benchmarking; human evaluation; RLHF-style preference grading; annotation QA; evaluator calibration; hallucination analysis; failure taxonomies; red teaming

Tool / Function Calling: tool selection; schema and argument validation; multi-step state transitions; retry/recovery paths; workflow and side-effect assertions

Automation & APIs: Cypress; Playwright; Pytest; Appium; Katalon; Postman/Newman; REST and contract testing; K6 performance testing

RAG & Grounding: retrieval relevance and coverage; context quality; source/citation fidelity; answer correctness; query routing; graph-based retrieval

Prompt & Quality Engineering: prompt engineering; prompt regression; instruction-following tests; adversarial cases; BDD/Gherkin; requirements traceability

Engineering & CI/CD: JavaScript; TypeScript; Python; Java; PowerShell; GitHub Actions; Jenkins; Docker; Jira; CodeQL; HTML reporting

PROFESSIONAL EXPERIENCE

AI Systems & QA Engineer (Automation Lead) - Opex Consulting - RegTech365

Lagos, Nigeria | December 2025 - Present

- Lead QA automation and quality engineering for RegTech365 products supporting regulated financial-services workflows, mapping risk-critical requirements to BDD scenarios, RTM identifiers and CI quality gates.
- Designed deterministic tests for LLM-backed and integration-heavy flows using API interception, static fixtures, JWT state injection, empirical selector governance and mocked webhook/WebSocket synchronization.
- Built Cypress/Playwright and API suites with Postman, Newman and PowerShell across onboarding, AI profile mapping, assessments, approvals, reporting and cross-product synchronization, surfacing high-severity blockers before release.
- Built a BDD-first automation framework that reduced test-automation delivery time by approximately 70%, and integrated AI-assisted failure triage, severity classification and CI-ready diagnostic reporting through TestAlgnite.
- Coordinated automation requirements, stable test hooks, test data, environment access and release evidence across product, engineering and QA stakeholders.

Software QA Engineer - Walure

Remote | December 2023 - December 2025

- Owned functional, regression, API and automation QA for high-traffic SaaS products, building Cypress and Katalon suites that increased automated test coverage to approximately 95%.
- Developed and executed Postman-based API and microservice tests, contributing to an estimated 60% reduction in high-severity pre-deployment blockers.
- Produced test plans, requirement traceability, release evidence and defect reports while collaborating with developers and product stakeholders through the full defect lifecycle.

Freelance / Contract QA Engineer - Testlio and independent clients

Remote | 2024 - 2025

- Delivered functional, regression, exploratory and mobile testing across web and mobile products for global engagements including BBC News, Matterport, FELLI Music and Wellcrafted.
- Executed cross-browser and device coverage, documented reproducible defects and supported release validation across distributed product teams.

SELECTED AI RELIABILITY & EVALUATION WORK

TestAlignite - AI-first QA diagnostics and CI reporting framework | [View project](#)

- Built a Node.js/Cypress reporting pipeline that merges test artifacts, isolates failures, applies model-assisted root-cause and risk analysis, and generates standalone HTML reports for CI review.
- Designed safe fallback behaviour when model credentials are unavailable, plus structured outputs for defect summaries, severity classification, remediation suggestions and regression intelligence.

LexTutor - Legal AI exam-intelligence and route-validation system | [View project](#)

- Designing a graph-structured legal intelligence pipeline with query clarification, topic detection, retrieval, context assembly, route-validated reasoning, weakness analytics and adaptive drill recommendations for Nigerian Bar Finals preparation.
- Created evaluation criteria covering authority accuracy, fact-rule pairing, command-verb fidelity, trap resistance, commitment discipline, hallucination control and examiner-facing quality; the structured system ranked first in the internal benchmark at 326/360.

Comparative Legal AI Benchmark - Five-system model evaluation and failure analysis

- Designed and ran a benchmark across five AI systems and eight civil-litigation tasks using a nine-dimensional, 45-point rubric; produced section-level diagnostics, system profiles and engineering controls from recurring failure patterns.
- Translated findings into validation rules for source-defined route exposition, authority-proposition locking, statutory element checks, factual-unit preservation and hallucination-resistant legal reasoning.

RegWatch Automation - Deterministic QA for AI-assisted regulatory workflows | [View project](#)

- Implemented BDD-first, RTM-aligned testing for compliance workflows, with deterministic mocks for AI requests, webhooks and real-time updates; added risk tiers and release-blocking CI gates for critical user journeys.

API Detective - Enterprise API and integration-testing framework | [View project](#)

- Built a repeatable integration-testing backbone using Postman scripts, Newman, PowerShell, GitHub Actions and Jenkins, with dynamic HTML reporting for high-stakes microservice workflows.

PUBLICATION

The Reliability Gap - Position paper on trustworthy legal AI | [View project](#)

- Authored a 16-page field analysis distinguishing fluent legal output from reliable legal reasoning and defining authority grounding, procedural fidelity, verified retrieval, commitment discipline, honest boundaries and traceability as core system requirements.

RECOGNITION

- Wema Bank Hackaholics 6.0 - 3rd-place regional qualifier (top 1% of 1,000+ applicants); contributed QA workflows that accelerated fintech prototype delivery.

EDUCATION & CERTIFICATIONS

Barrister at Law (BL) - Nigerian Law School, 2025

Bachelor of Laws (LL.B) - Redeemer's University, 2022

Certifications: JavaScript (Meta); API Automation with Postman; Software QA Engineering (Walure); Information Security and API Security Fundamentals.