

Safwan Saba

AI Test Lead | Quality Engineering Architect | LLM & ML Evaluation | Test Platform Engineering

(352) 727-9967 | safwan@safwansaba.info | [Portfolio](#) | [linkedin](#) | [GitHub](#)

AI Test Lead and Quality Engineering Architect with 14+ years of experience building quality strategy, automation platforms, and release gates for enterprise systems in finance, healthcare, and compliance. Hands-on in testing AI/ML and LLM applications: designing evaluation datasets and scoring rubrics, measuring accuracy, hallucination, and retrieval quality, and wiring continuous evaluation into CI/CD. Led teams of nine engineers across three offshore teams, and set the standards for reliable, trustworthy delivery. Creator of Sentinel, an AI-assisted self-healing QA platform, and an LLM-powered resume assistant validated with a structured evaluation suite.

Notable recognition: Discovered and responsibly disclosed a critical iOS security vulnerability ([iOS 5.0.1](#)) affecting 8M+ users worldwide.

Core Competencies

- **AI/ML Testing & Evaluation:** AI test strategy, LLM evaluation, RAG evaluation, retrieval accuracy, hallucination and groundedness detection, prompt regression testing, evaluation datasets, scoring rubrics, accuracy, precision/recall/F1, robustness and adversarial testing, bias and fairness checks, drift monitoring, guardrail validation, human-in-the-loop review, agent workflow testing, Gemini API, OpenAI API
- **Quality Leadership:** Test strategy and ownership, quality gates, acceptance criteria for AI features, release-readiness reporting, risk-based testing, team leadership and mentoring, cross-functional collaboration with data science, engineering, and product
- **Automation & Frameworks:** Playwright (TypeScript, Java), Selenium WebDriver, Cypress, Appium, REST Assured, Cucumber/Gherkin, SpecFlow, TestNG, JUnit, Page Object Model, fixtures, network mocking, parallel execution, flaky-test remediation
- **Languages:** Python, Java, TypeScript, C#
- **CI/CD & MLOps-Adjacent:** Azure DevOps, GitHub Actions, Jenkins, Docker, Git, Maven, evaluation runs as pipeline quality gates
- **Distributed Systems Testing:** Microservices, contract testing (Pact, consumer-driven contracts), gRPC, event-driven and asynchronous workflows, idempotency, retries, partial-failure recovery
- **Data, Cloud & Observability:** MySQL, PostgreSQL, MongoDB, JDBC, AWS (EC2, S3, Lambda), Azure, Splunk, Grafana, Allure, Jira Xray

Professional Experience

Senior QA Consultant | Asman Software Solutions | Tampa, FL | Jun 2025 – Present

- Define AI testing strategy for client AI-enabled applications, covering functional, integration, and user-facing behavior, with entry and exit criteria for each release stage.
- Design evaluation datasets and scoring rubrics for LLM and RAG features, measuring answer accuracy, groundedness, hallucination rate, and retrieval relevance, and track results release over release.
- Run prompt and model-version regression testing, with evaluation runs gated in CI so quality regressions are caught before release.
- Validate guardrails, adversarial and edge-case prompts, and human-in-the-loop review workflows for AI features.
- Build Playwright and Python automation with reusable fixtures, API setup, network mocking, parallel execution, trace analysis, and deterministic test data.
- Established consumer-driven contract testing with Pact to catch provider incompatibilities early and reduce late integration failures.
- Use AI-assisted techniques for failure analysis and scenario generation, improving defect investigation speed and automation stability.

Lead Quality Engineer | PwC | Tampa, FL | May 2022 – May 2025

- Led and mentored nine engineers across three offshore teams, setting code-review standards, framework-design patterns, and testability practices that contributed to a 25% reduction in critical production defects.
- Owned quality strategy for three enterprise applications: automation architecture, risk-based test strategy, release-readiness criteria, and quality reporting across UI, API, integration, and database layers.
- Tested AI/ML-enabled enterprise applications hands-on, validating model outputs for accuracy, precision/recall/F1, and consistency against curated evaluation datasets and defined acceptance thresholds.
- Designed evaluation datasets and scoring rubrics for ML and LLM-driven features, covering normal, edge-case, and adversarial inputs, and tracked results release over release to detect regressions.

- Built Python and Playwright automation to exercise AI-backed workflows end to end (UI, API, and data layers), verifying model responses, fallback behavior, and error handling under partial failure.
- Performed robustness, bias, and drift checks on model outputs across data segments and model versions, and escalated findings to data science and product with reproducible evidence.
- Integrated model evaluation runs into Azure DevOps pipelines as quality gates, blocking releases when accuracy or regression thresholds were not met.
- Established human-in-the-loop review steps for low-confidence or high-risk AI outputs and defined release-readiness criteria for AI features in a regulated, compliance-sensitive environment.
- Partnered with data scientists and engineers to define testability requirements for models, including versioned test data, reproducible evaluation runs, and clear pass/fail criteria.
- Architected a Playwright and TypeScript platform spanning UI and API testing, increasing automated coverage by 45%.
- Designed a scalable Java, Cucumber, and BDD framework that cut regression execution from eight hours to under three, saving about 20 engineering hours per sprint.
- Integrated 200+ automated API and database checks into Azure DevOps pipelines as continuous-testing quality gates, reaching 90% automation of eligible regression scenarios in a SAFe environment.
- Built reusable API and database-validation components with REST Assured, JDBC, and MySQL, improving automation performance by 40%.

Quality Engineer | Mednax | Sunrise, FL | Aug 2021 – May 2022

- Architected a modular Page Object Model framework with Playwright (Java and TypeScript) for critical healthcare applications.
- Increased test coverage by 35% and eliminated 40% of manual regression effort through targeted automation.
- Implemented an API testing strategy for healthcare data integration pipelines using REST Assured.

Quality Analyst | PwC | Tampa, FL | Nov 2020 – Aug 2021

- Performed end-to-end testing across front-end, back-end, API, and database layers.
- Developed C# and Selenium automation with SpecFlow BDD in Azure DevOps.

Quality Analyst | Fannie Mae | Tysons, VA | Jul 2020 – Nov 2020

- Built automated regression suites with Java, Selenium, Cucumber, and Jenkins, reducing manual testing effort by 30%.
- Validated RESTful services with Postman and REST Assured; performed Appium mobile automation on iOS and Android.

Earlier Career (2012 – 2020): SDET and Quality Engineering roles at Mobius Logic (Tysons, VA), Responsible Robotics (Tysons, VA), and MBC Group (Dubai, UAE). Built hybrid Selenium frameworks, configured Jenkins CI pipelines, and tested enterprise-scale media streaming platforms.

AI/ML Testing Projects

Ask AI About Me: LLM-Powered Interactive Resume | [Live Demo](#)

- Built an interactive AI chatbot governed by system prompt engineering and Next.js context injection to accurately guide Gemini 2.0 based on my professional background and skills

Sentinel: Autonomous Self-Healing QA Platform | [Live Demo](#)

- Architected a full-stack autonomous QA platform with AI-powered self-healing using Gemini 2.0 reducing manual locator maintenance by 80%
- Engineered multimodal failure analysis pipeline ingesting DOM snapshots, screenshots, and stack traces for 90%+ fix accuracy
- Built real-time defect detection feed with 3-second polling, live mission telemetry, and an integrated Community Hub for collaborative triage
- Implemented pre-flight build validation (mvn test-compile) to ensure AI-generated code never breaks the project

Neural Profile: Interactive Career Visualization Platform | [Live Demo](#)

- Built an interactive neural network visualization of 15 years of career experience, where each skill is a node and connections represent relationships between technologies, projects, and expertise .

Education & Certifications

- **B.S., Software Engineering and Automated Systems**, State Engineering University of Armenia, 2004 – 2008
- [Supervised Machine Learning: Regression and Classification](#), Stanford University | DeepLearning.AI
- [ISTQB Certified Tester, Foundation Level \(CTFL\)](#), ASTQB