ANKIT KALLURAYA

AI QUALITY ENGINEERING | TEST AUTOMATION | APPLIED AI

Australia | kallurayaankit@gmail.com
ankitkalluraya.is-a.dev | github.com/kallurayaankit

PROFILE

Quality engineer and independent AI consultant combining regulated software testing with hands-on automation and applied AI development. Builds test frameworks, evaluates LLM outputs, and turns requirements into reproducible evidence. Experience spans functional, regression, compliance, API and performance testing, supported by public projects in AI-assisted testing and browser agents.

PROFESSIONAL EXPERIENCE

Independent AI Quality Consultant & Founder

Independent | Full-time | Australia, Remote | Aug 2025 - Present

- Built a Doer/Adversary/Judge framework to detect prompt injections and hallucinations, validating outputs from GPT-4, Claude 3.5 Sonnet and Llama 3.
- Engineered an AI Agent Validation framework using LLM-as-Judge to establish quality baselines and identify reliability issues before production.
- Architected a mock Model Context Protocol (MCP) server with GitHub Actions CI/CD integration; published portfolio code demonstrates infrastructure ownership and defect prevention.
- Delivered a Docker and Locust performance suite with enforced p95/SLA thresholds, helping early-stage teams validate scalability without dedicated DevOps.
- Partnered with developers to integrate testing into shipping cycles, shorten feedback loops and support faster releases.

QA Test Analyst / Senior Test Engineer / Project Lead

iTech Labs (GLI Subsidiary) & Gaming Labs International
Australia | Nov 2021 - Aug 2025

- Designed and executed system test plans for online gaming and wagering platforms, translating jurisdiction-specific regulatory requirements into test scope.
- Delivered functional, regression and compliance testing; documented and analysed defects in JIRA and coordinated resolution with global teams.
- Managed end-to-end certification work and prepared reports for clients and regulators; introduced templates to improve consistency and reduce manual errors.

EDUCATION

Master of Electrical Engineering with Business (Research)

The University of Melbourne | Feb 2017 - Jan 2021

SELECTED WORK & TECHNICAL CAPABILITIES

FLAGSHIP PROJECTS

ReleaseRadar | AI-assisted regression testing

JavaScript, OpenAI, Playwright, GitHub Actions
github.com/kallurayaankit/release-radar

- Demonstrates how a passing checkout test suite can miss defects. Freezes assertions, challenges three seeded bugs, reviews test additions and reruns unchanged assertions against a corrected build.
- Exports reproducible test evidence and Playwright specifications. Hosted prepared/imported cases work within one checkout requirement; live test generation requires a configured local server.

Jev Flight Agent | Standalone AI application

Python, Jev / TypeSafe AI, Playwright
github.com/kallurayaankit/jev-flight-agent

- Turns natural-language flight-search goals into typed browser operations. Jev selects observed DOM targets; a separate language model generates field text and Playwright executes actions.
- Checks stale page state, bounds execution and records outcome evidence. Includes demonstrations and offline tests; local setup requires API keys and Chrome. Search prototype, not a booking service.

ADDITIONAL ENGINEERING WORK

Assay - Calibrates LLM judges against human-labeled examples using Cohen's kappa; gates scoring and retains calibration provenance and model-call budgets.

Jev Ticket Router - FastAPI prototype for team, urgency and sentiment classification; routes low-confidence department decisions to human review.

Playwright Test Automation Framework - Python/Pytest UI automation with page objects, JSON test data, parallel execution and GitHub Actions.

TECHNICAL SKILLS

Languages & development: Python, JavaScript, Java; FastAPI, Flask; Git

Quality engineering: Playwright, Pytest, Selenium; API testing with Requests and JSON Schema; functional, regression, negative and contract testing

AI quality & applications: LLM-as-Judge, adversarial evaluation, prompt injection, hallucination checks, mock MCP, typed agent actions, Ollama

Performance, data & delivery: Locust, p95/SLA validation, MongoDB, pandas, GitHub Actions, Docker, JIRA, Confluence, Postman

Delivery & collaboration: Requirements analysis, defect triage, certification reporting, regulatory test scoping, cross-team coordination