

Raghav Gupta

SDET WHO EXTENDED TEST-DESIGN DISCIPLINE FROM DETERMINISTIC SYSTEMS TO AI AGENTS

raghavg2705@gmail.com [+91 98184 76527](tel:+919818476527) in/raghavg27 github.com/raghavg27 Delhi-NCR, India

PROFILE

Five years proving out other people's systems – mobile, web, warehouse robotics – then AI agents arrived and broke the assumption test engineering was built on: that correctness is binary. At Handshake AI I design reproducible evaluation environments for coding agents; on my own time I build the agent systems themselves, the way a tester would. Looking for the role where those two halves are the same job.

WHAT I BRING

- Test design that survives non-determinism – objective pass/fail criteria for AI agents whose correctness is probabilistic.
- Owning test infrastructure end to end: in-house PyTest + Page Object Model frameworks, CI/CD-integrated, flake-free.
- Builds the systems he evaluates – multi-agent LLM applications designed around grounding, validation and citation.
- Full-stack quality range: functional, API, mobile, and JMeter performance/load testing under peak throughput.

EXPERIENCE

Handshake AI

MAR 2026 – PRESENT · REMOTE

QA ENGINEER – AI AGENT EVALUATION (CONTRACT)

- Designed end-to-end test scenarios for **AI coding agents** – each packaged as a reproducible **Dockerised environment** with a written specification, a reference implementation, and an automated verification suite defining objective pass/fail criteria.
- Applied **boundary, negative and exploratory testing** to surface agent failure modes in long-horizon planning, environment state handling, and recovery from failed tool calls.
- Enforced **determinism** and eliminated flaky verification so results measure agent capability rather than environment defects.
- Extended five years of test-design discipline to **non-deterministic systems**, where correctness is probabilistic and assertions must tolerate valid variation in approach while still catching wrong outcomes.

Grey Orange

SEP 2025 – DEC 2025 · GURUGRAM, INDIA

SOFTWARE ENGINEER – QA / SDET

- Owned end-to-end sprint QA – planning, test design, automation, defect tracking and retrospectives – landing **on-time releases** with improved defect containment.
- Built automated regression suites for warehouse and robot workflows in **PyTest, Selenium, Playwright and Page Object Model**, cutting manual effort **60–70%**.
- Ran performance and load testing for high-throughput bulk operations with **JMeter** and REST API testing, validating SLA compliance under peak load.
- Published pipeline metrics to **Grafana** for coverage, failure-trend and release-readiness visibility across stakeholders.
- Triaged nightly automated pipelines – isolating flaky tests and coordinating fixes via JIRA – reducing false positives **25%**.

Shway

JUN 2023 – SEP 2025 · REMOTE

SOFTWARE ENGINEER – QUALITY ASSURANCE

- Led functional, system and regression testing across **iOS and Android** using manual QA plus Selenium/Appium – **zero critical production defects** across releases.
- Ran end-to-end API testing with Postman, RestAssured, Requests and Swagger, cutting backend defect leakage **40%**.
- Executed load testing for high-traffic flows with JMeter, improving response times under peak load **30%**.
- Designed an in-house data-driven automation framework (PyTest, Page Object Model) that reduced test maintenance **35%**.
- Integrated test execution into CI/CD with **Jenkins, Git, Docker and Kubernetes** for quality checks on every build.

Yellow.ai

JUL 2021 – JUN 2023 · REMOTE

CUSTOMER SUCCESS ENGINEER

- Acted as **quality gatekeeper** for enterprise clients – owning pre-release validation, defect triage and go-live sign-offs for conversational workflows.
- Triaged incidents, set severity, and verified fixes in staging and production, lifting CSAT and reducing repeat incidents.
- Led technical UAT sessions with enterprise customers, converting business feedback into product improvements.

SELECTED PROJECTS

squared-up INDIA-FIRST EXPENSE SPLITTER · DJANGO/DRF + REACT PWA [github.com/raghavg27/squared-up](#) ↗

Money core held in integer paise and verified with Hypothesis property-based tests; UPI one-tap settlement, installable PWA, Dockerised.

equity-crew MULTI-AGENT EQUITY RESEARCH · CREWAI [github.com/raghavg27/equity-crew](#) ↗

Five specialised agents fusing fundamentals, neural news search, from-scratch technical indicators and sector peer benchmarking into a validated BUY/HOLD/SELL PDF report.

git-guide AGENTIC RAG OVER GITLAB DOCS [github.com/raghavg27/git-guide](#) ↗

CrewAI + ChromaDB + Streamlit; async parallel processing and smart routing, with every answer grounded and citation-enforced against source docs.

Steno WRITING STANDARD FOR AI PROSE [steno-ai.vercel.app](#) ↗

A tiered editing standard that makes AI-generated prose read as though a person wrote it. Ships as a Claude skill and a portable prompt.

TOOLKIT

LANGUAGES & DATA

Python · JavaScript · SQL · PostgreSQL · MongoDB / NoSQL · Bash

AI / BACKEND

CrewAI · Agentic RAG · ChromaDB · Django / DRF · React · Node.js · REST / OpenAPI

TEST & AUTOMATION

Selenium · Playwright · Appium · PyTest · Hypothesis · JUnit · Page Object Model · JMeter · Postman / RestAssured

DEVOPS & CI/CD

Jenkins · Docker · Kubernetes · Git / GitHub Actions · Grafana · JIRA · TestRail · Confluence

EDUCATION

DEGREE B.E. Computer Engineering · 2017–2021
SRM Institute of Science and Technology

AVAILABILITY

ENGAGEMENT	Independent contractor / B2B · EOR · full-time via an India-registered entity
LOCATION	Delhi, India · remote worldwide, any timezone
WORK AUTHORISATION	Indian citizen. No visa or sponsorship required for contract engagements.