NIRAJ PAUDEL

SDET and Test Automation Engineer | 10 years in software engineering | Testing AI Systems

Auckland | New Zealand resident, full work rights | itsmenirajpaudel@gmail.com
nirajpaudel.com | linkedin.com/in/itsmenirajpaudel | github.com/NPZlatu

SDET
test automation on Monotype's myfonts.com

AI-native
builds and tests AI systems with Claude Code agents, daily

20 minutes
regression run in CI, down from 1 to 2 days by hand

Master of IT
Auckland, with research on LLM evaluation

PERSONAL STATEMENT

I am a software engineer with ten years of building products, two of them as a full-time Software Development Engineer in Test. On Monotype's myfonts.com I wrote BDD browser tests in Protractor, Selenium and Cucumber, ran them in Jenkins, and cut the regression cycle from one to two days of manual checking to a twenty minute automated run.

At Nylon, an AI company, I work as a software engineer on the knowledge base pipeline. My work is half building and half verification, both with AI: scraping and ingesting legal data on Node.js and AWS, then checking that what was built and loaded is correct. My Master's research measured how far automated LLM evaluation can be trusted.

As AI writes more of the code, the harder and more valuable work is proving that what it built is correct. That is the work I want to do full time, so I am looking for a test automation or quality engineering role.

KEY SKILLS

Test automation	Playwright, Puppeteer, Selenium, Protractor, Cucumber (BDD), UI and REST API tests, Postman
Testing	Functional, regression, exploratory and end to end testing, test planning and strategy, test case design, defect reporting
Testing AI systems	LLM evaluation with RAGAS, RAG data validation, MCP tool testing across Claude, ChatGPT and Gemini
Data validation	SQL on PostgreSQL, Qdrant vector database (search, payload and parity checks), source reconciliation
CI/CD	GitHub Actions, Jenkins, Docker, Terraform
AI-native workflow	Claude Code and parallel AI agents, MCP servers, agent and CodeRabbit code review, human-verified releases
Languages and runtime	TypeScript, JavaScript, Node.js, Python, C#, SQL
Cloud and monitoring	AWS (Lambda, SQS, S3, ECS), Datadog, CloudWatch
Tools and process	Vitest, Jest, Mocha, JMeter (performance testing), TestRail, Jira, Git, Agile and Scrum

ENGINEERING AND TEST AUTOMATION EXPERIENCE

Software Engineer (Data)

May 2025 to present

Nylon (formerly Law Cyborg), Auckland

AI research tools for tax and law professionals. Joined as an intern, kept on part-time, then full-time.

- **Knowledge base pipeline.** My main work: built scrapers and ingestion for legal sources across New Zealand, Australia and the UK on Node.js, AWS Lambda and SQS, loading a Qdrant vector database. Includes Puppeteer scrapers for sites behind Cloudflare and Akamai.
- **Weekly scraper runs.** Ran the weekly scraper across the legal sources on staging and production, reviewed what changed, and manually verified the results before they reached production.
- **Data validation.** Checked each new source against the original website: record counts, broken links, missing and duplicate documents, and whether a search returns the right text. Wrote the staging to production sync for Qdrant with a dry run and an ID-by-ID parity check.

- **Monitoring.** Set up Datadog monitoring for the pipeline, showing what each step is processing and where a run fails, and traced failed runs to their cause.
- **MCP server and end to end tests.** Built Nylon's MCP server, which puts Nylon's legal search inside Claude, ChatGPT and Gemini, and tested the sign-in (OAuth 2.1) and every tool against each AI client. Added Playwright end to end tests, written with Claude Code, to the web app's GitHub Actions pipeline.
- **AI-native workflow.** As a team we built our AI workflows and work in them every day: Claude Code agents write the code and tests, a review agent challenges each change, and the engineer verifies the result on staging. Releases are gated by the team's LLM benchmark (gold question sets, an independent LLM judge).

Software Development Engineer in Test (SDET) May 2019 to Jan 2021

UBA Solutions, Lalitpur, Nepal (client: Monotype, USA)

Full-time test automation on myfonts.com, Monotype's online font store.

- **Automation.** Wrote automated browser and API tests in Protractor, Selenium and Cucumber for the main site, its micro-sites and its microservices, turning acceptance criteria into test scenarios the whole team could read.
- **CI and performance.** Ran the suites in Jenkins on every build and investigated each failure. Load tested services with JMeter.
- **Functional and regression.** Tested new features by hand before automating them, and cut the regression cycle from one to two days of manual checking to a twenty minute automated run.
- **Bug reports and test management.** Reported bugs with clear steps to reproduce, retested each fix, and kept test cases and runs up to date in TestRail.

EARLIER SOFTWARE ENGINEERING EXPERIENCE

Senior Software Engineer and Team Lead Dec 2022 to Mar 2024

UBA Solutions, Lalitpur, Nepal (client: Monotype, USA)

- Returned to the myfonts.com team as a developer. Moved the site from Vue 2 to Vue 3 and built the Bundles feature, with tests for each release. Reviewed code and coached less experienced teammates.

Senior Software Engineer and CTO Jan 2021 to Nov 2022

Business Pazes (start-up), Kathmandu, Nepal

- Ran the technical side of a start-up: chose the tech stack, built the product hands-on with a small team, and set up code review and testing before every release.

Software Developer 2015 to 2019

B.M. Nepal and Spiralogics, Kathmandu, Nepal

- Built modules for the Bookmundi travel booking site, payment gateway integrations, and web applications on the MEAN stack.

RESEARCH AND PROJECTS

Reliability of RAGAS Metrics across LLM Judges (Master's research, 2026)

Can one LLM be trusted to judge another LLM's answers? Three LLM judges (GPT-4o mini, Gemini 2.5 Flash-Lite, Llama 3.1 8B) scored the same RAG answers with RAGAS, 1,800 scores in all. Result: reliable for comparing systems, not for exact scores. The judges ranked answers in the same order (Spearman 0.82 to 0.89) but scores moved by about 0.1 between judges, and one metric, Answer Relevancy, could not detect hallucinated answers.

NZ Visa Application Checklist (nirajpaudel.com)

An interactive checklist I built after my own visa application. Other applicants now use it to prepare their documents.

EDUCATION

Master of Information Technology, Auckland Institute of Studies 2024 to 2026

Certificate of Academic Excellence, Postgraduate High Achievement (2025)

Bachelor of Electronics and Communication Engineering, Tribhuvan University, Nepal 2010 to 2014

REFEREES

Available on request.