Shawn Skelly

Software Development Engineer in Test · Test Automation & Tooling
hello@shawnskelly.com · Arlington, VA · shawnskelly.com · github.com/spskelly

PROFESSIONAL SUMMARY

Software and test-automation engineer who designs frameworks and tooling for reusability, scale, and reliability: CI/CD-integrated test frameworks, automation across multiple business domains, and, more recently, evaluation and retrieval tooling for AI agents. Strong track record on federally deployed platforms: cutting regression execution time dramatically, raising the engineering standards of inherited test suites, and reaching past the standard QA remit to diagnose and prototype fixes for deeper system problems. A recurring operating pattern runs through the work: opinionated, standards-driven taxonomies feeding deterministic scoring, with LLMs positioned as auditors rather than decision-makers.

WORK EXPERIENCE

IBM (formerly Octo Consulting) Jul 2022 – Present

Assoc. Software QA Engineer / QA Lead *Jan 2024 – Present*

Test-and-evaluation engineering for a federally deployed Salesforce (GSA) platform, covering test automation, internal tooling, and AI agent evaluation.

- Diagnosed retrieval-quality failures in a federally deployed Salesforce knowledge search system and built an Apex-native proof-of-concept using hybrid keyword + vector retrieval via reciprocal rank fusion (RRF) with LLM-summarized results and no Agentforce dependency, matching the synthesized-answer UX at ~100× lower per-search cost (≈$0.0009 vs $0.10/action).
- Built a Python evaluation harness for the Agentforce knowledge-summary agent: deterministic scoring over a four-tier question taxonomy, with an HTML report of hit@1/3/5, gold-article rank, and confuser rank by question type, enabling regression-style measurement across model and configuration changes.
- Replaced an inherited Katalon UI-regression suite with a hand-written scheduleable Apex test framework designed for reusability, scalability, and reliability, cutting regression execution time ~95% and remaining UI-suite runtime ~50%.
- Built and own end-to-end test automation (UI, Apex, and Postman API suites) and a CI/CD pipeline across three Salesforce business domains as the sole engineer responsible for automation.
- Built an internal sprint-grooming tool that ingests a full Jira sprint plan and scores each story's clarity, point estimate, and subtask breakdown against a multi-page standards-and-domain system prompt I authored, emitting a single HTML report with gap identification and actionable refinement suggestions.
- Delivered internal enablement on using AI as a QA lead and a Center of Excellence session on UAT methodology; routinely facilitate stakeholder requirements discussions and unblock cross-team delivery.

Assoc. Software QA Engineer *Jul 2022 – Jul 2023*

Automated testing for the U.S. Department of Labor (OSHA).

- Cut test execution time >50% and raised throughput ~70% by optimizing a Selenium framework for parallel execution, with Jenkins CI/CD for automated runs and reporting.
- Integrated automated Section 508 accessibility testing into CI, catching hundreds of compliance issues earlier in the lifecycle.

Alliance Tek Solutions Feb 2017 – Jun 2022

Manual Tester → QA Automation Tester → QA Automation Engineer

- Built a Selenium/Java BDD automation framework (TestNG, Maven, Cucumber, Selenium Grid on Docker) for cross-browser functional, regression, and smoke testing, with data-driven cases via Apache POI and defect tracking in Jira and Azure DevOps.

PROJECTS

TALON: AI-Powered Land-Intelligence Platform *(Python, FastAPI, PostGIS, Astro/Preact, Proxmox)*

Full-stack geospatial platform for rural-property analysis, owned end-to-end from app to self-hosted bare metal. Staging deployed; preparing open beta with 14 western-NC counties at 1m DEM (depth over breadth).

- Full-stack app: FastAPI/SQLModel + PostGIS backend and Astro/Preact frontend, with natural-language parcel search compiled to SQL, AI-generated parcel summaries, and interactive chat over parcel data.
- Geospatial analysis: a LiDAR pipeline for individual-tree detection and timber volume, plus layer-driven microhydro and soil-suitability (SSURGO) analysis and a 3D LiDAR terrain viewer.
- Infrastructure: self-managed Proxmox on a bare-metal cloud server, handling VM/container provisioning, reverse-proxy and networking scaffolding, and the full deployment substrate, run solo.
- Engineering: hexagonal (ports-and-adapters) architecture, main/staging/develop branching with CI quality gates, and 50+ Alembic migrations.

RepoAudit: Code-Intelligence & AI-Maturity Analyzer *(Python, GitHub Actions, OWASP/CWE, OSV)*

Repository analysis platform with six profiles: security (OWASP/CWE, OSV CVE scanning), architecture, documentation, portfolio, diagrams, and an original tri-axis AI-maturity model (cognitive, operational, agentic). Rubric-based deterministic scoring, with an LLM validation pass confirming the higher maturity levels; runs as a CLI or zero-install GitHub Actions mode.

VoiceForge: LLM Fine-Tuning & Corpus-Engineering Pipeline *(Python, Llama-3.1, QLoRA/Unsloth, Ollama)*

End-to-end ETL and fine-tuning pipeline that normalizes heterogeneous personal text (chat exports, Reddit dumps) into a provider-neutral, authorship-tagged corpus, then fine-tunes a local model on it. QLoRA fine-tune of Llama-3.1-8B on a single RTX 3090 (Unsloth/WSL2), multi-task with response-only loss masking; root-caused a non-obvious CUDA OOM to a library gradient-offload bug. The hard part is data quality: a paste-vs-voice stage separates real writing from pasted code/quotes, and intent classifiers fixed a bias that ended every generation in a question. Deployed locally via Ollama (GGUF, q4_K_M); 56 tests.

SKILLS

Languages: Python, Java, Apex, JavaScript, SQL, Groovy

AI & ML: RAG, hybrid retrieval (keyword + vector, RRF), embeddings, LLM fine-tuning (QLoRA/PEFT), evaluation harnesses, deterministic scoring, LLM-assisted labeling, trained ML classifiers, GPU/CUDA, multi-provider (OpenAI / Anthropic)

Test & Automation: Apex test development, Selenium, Postman / API testing, Katalon, BDD/TDD, CI/CD (Jenkins, GitHub Actions)

Geospatial & Data: PostGIS, PostgreSQL, LiDAR processing, spatial / hydrological modeling, ETL, data analysis

Platforms & Standards: Salesforce (Apex / LWC), Docker, Linux, Git; NIST 800-53, OSCAL, Section 508

EDUCATION

Western Governors University · BS, Software Engineering In progress (Expected 2027)