Uliana-Olena Sakhno

AI Quality Engineer — LLM Agents · Evals · Guardrails

Remote · Ukraine (CET+1) · Open to B2B contract · uliana@sakhno.tech · +38 (097) 440-22-05
linkedin.com/in/ulianasakhnoqa · github.com/VelenShell · olena.sakhno.tech

SUMMARY

QA engineer with 6+ years in backend, data and security testing, now building and validating LLM agents in production. On a 3-person AI team serving about 100 users, I build agents on Claude, set up PHI/PII guardrails for HIPAA, and verify agent output before it reaches developers or production data.

My edge is knowing where AI fails. I design the logic, let coding agents implement it, and review the code and results line by line. Background in regulated healthcare, endpoint security and a merge of two legacy databases with 375 and 582 tables.

SKILLS

LLM & agents	Anthropic Claude on Amazon Bedrock · LiteLLM gateway · MCP and UTCP · Claude Code and Claude Desktop rollout · agentic workflows · prompt and context design
Guardrails & evals	Microsoft Presidio via LiteLLM · PHI/PII detection and masking (HIPAA) · golden sets · multi-source validation · false-positive tuning
Data quality	SQL (strong) · PostgreSQL · MS SQL Server · BigQuery · Snowflake · MongoDB · Redis · Elasticsearch · Great Expectations · migration reconciliation
Code	Python and TypeScript: code reading and review, AI-assisted implementation with Claude Code (daily) · Streamlit · GitHub Actions
Testing	API testing (Postman, Postman Flows, Fiddler, REST) · integration and backend · root-cause analysis · compliance and security testing · Jira, TestRail

EXPERIENCE

AI Engineer — Agiliway Mar 2026 – Present

Official title Senior QA Engineer. One of 3 engineers on the AI team of a client project with 84 engineers and a 20-year-old codebase. Stack: Claude on Amazon Bedrock and Claude subscriptions behind a LiteLLM gateway, Presidio, MCP/UTCP.

- Built an agentic data-migration loop: Python scripts profile the legacy database and pass structured data templates to an agent, which has inferred 826 undocumented business rules so far (e.g. a record cannot be updated before it was created).
- Built a validation step for agent-proposed rules: each rule is checked in parallel against Jira, Confluence and the code, needs confirmation from at least 2 sources, and is flagged if the evidence is outdated. Clear cases are auto-approved; about 20–25 went to manual review and 14 were rejected.
- Generated 12,071 checks for the first 4 of 16 migration domains (11,764 from the schema, 307 from business rules; 1,544 blocking integrity checks, 10,527 quality checks) and used the failing checks to clean the target: unified formats that changed across 20 years of data, so two inconsistent databases became one clean dataset.
- Mapped a merge of two legacy databases (375 and 582 tables, 222 shared) column by column. Started as my proof of concept, a Python mapping tool that produced a full match report in about 5 minutes instead of a week, which grew into a cross-team initiative. In that POC I caught columns narrowed from VARCHAR(64) to VARCHAR(32) that silently truncated Chinese text.
- Built an agent that watches schema changes in the legacy database and drafts the matching DTOs; it runs about weekly and its drafts have been accepted without changes so far.
- Set up and tuned Presidio PHI/PII guardrails on the LiteLLM gateway for HIPAA: tested with synthetic records built from real data templates and cut the guardrail set after finding over-masking that redacted ordinary words as names. Prompt storage is off by design.

- Helped roll out the gateway to about 100 users, including Claude Code and Claude Desktop (with a helper script), supported adoption, and wrote the privacy explanation that addressed fears of being monitored.
- Co-proposed moving engineers to Claude subscriptions routed through the same gateway when the client's Bedrock budget proved too small for everyday use.
- Migrating my agents from MCP to UTCP after the client restricted MCP; building an API test-automation framework as the base for agents that help manual QA write automated tests.

QA Engineer, Backend & Data — AllStars-IT May 2024 – Dec 2025

Project: Morphisec, an endpoint protection (cybersecurity) platform.

- Validated event ingestion, data consistency and reporting accuracy across Redis, Snowflake, PostgreSQL and Elasticsearch; traced recurring defects to their source through logs and multi-database queries.
- Ran security-oriented testing: surfaced potential vulnerabilities and validated protection mechanisms.
- Cut test-account setup from up to an hour to about 5 minutes with a Postman Flow that assembles the configuration JSON from selectable cases and creates the account in one run.

QA Engineer, Backend & Integration — TechMagic Jan 2023 – May 2024

Project: Huckleberry, a baby-care app (Android, iOS, web) used by 5M+ families, with a US product team.

- Tested backend APIs, data consistency and server logic; Stripe and Amplitude integrations; the move to a serverless backend.
- Proved that the A/B-testing tool the team had chosen reported unreliable results, took it to a call with the vendor, and the team switched to a more accurate tool.

QA Engineer — Cosm.care May 2022 – Dec 2022

- Compliance testing for a Canadian digital-healthcare product: data security, patient privacy and workflow correctness.

QA Engineer — Stepico Games Feb 2020 – May 2022

- Full QA cycle across a web gambling product, a Windows desktop racing game and mobile games.

PROJECTS

- **QA Context Radar** (Python, Streamlit, Claude API). Scores a ticket's readiness for testing; a golden-set eval measures where the model is grounded and where it invents.
- **Joblin** (Python, GitHub Actions, Claude API). LLM job-search pipeline with guardrails from its real failures: verifies each company's hiring market, bans overclaiming phrases, flags unknown data instead of guessing.
- **F1 2026 forecast** (Python, PostgreSQL, scikit-learn). ETL of 75M telemetry rows, hypothesis testing and a Monte Carlo simulation.

EDUCATION & LANGUAGES

- B.Sc., Automation & Computer-Integrated Technologies — Lviv Polytechnic National University, 2017–2022
- Data Analytics program, DAN.IT Education (208 hours), 2026 · TestGorilla QA / Test Engineer assessment: 94th percentile
- English: fluent (B2+/C1) · Ukrainian: native