



EN/JP Software Engineer in Test / QA Engineer (Agentic AI)

Job Information

Temp Agency

Randstad K.K., Professionals

Job ID

1600816

Industry

Other (Banking and Financial Services)

Company Type

Large Company (more than 300 employees) - International Company

Non-Japanese Ratio

Majority Non-Japanese

Job Type

Contract

Location

Tokyo - 23 Wards

Salary

6 million yen ~ 9 million yen

Refreshed

July 6th, 2026 18:41

General Requirements

Minimum Experience Level

Over 3 years

Career Level

Mid Career

Minimum English Level

Fluent (Amount Used: English usage about 50%)

Minimum Japanese Level

Business Level

Minimum Education Level

Bachelor's Degree

Visa Status

No permission to work in Japan required

Job Description

This is a rolling renewable contract role at an established foreign insurance firm.

Our client is establishing a brand-new capability within their AI Platforms Team and are looking for a **bilingual AI QA / Evaluation Engineer** to build evaluation foundation from scratch.

You will serve as the quality guardian for applications powered by LLMs, RAG pipelines, OCR, tool calling, and automated workflows. This role is perfect for a **technical QA professional** who is curious about LLM behaviors, proactive about finding edge-case risks, and passionate about continuous automated validation.

Key Responsibilities:

- **Build AI Evaluation Datasets**: Create and maintain robust "golden datasets" and test cases that act as unit tests for prompts, RAG responses, OCR outputs, and agent workflows.
 - **Test System Correctness**: Validate that agentic systems demonstrate sound reasoning, strict retrieval accuracy, correct tool execution, and reliable fallback handling.
 - **Mitigate Risk & Hallucinations**: Define methodologies to detect production hallucinations, measure answer quality trends, and build operational dashboards (Power BI, Grafana, etc.).
 - **Automate Evaluation Pipelines**: Build reusable test harnesses and integrate automated evaluation/regression checks directly into CI/CD release gates.
 - **Responsible AI & Compliance**: Evaluate applications for potential bias, unfair treatment, compliance issues, and data privacy risks within regulated environments.
 - **Cross-Functional Partnering**: Translate business expectations into measurable technical criteria alongside engineers, data scientists, and business stakeholders.
-

Required Skills

■Qualifications

- Min of 4- 5 years of work experience
 - Core experience in **Software QA, test automation, data validation, or AI/ML application quality engineering**.
 - Solid conceptual understanding of **LLMs, prompt engineering, RAG, OCR, tool calling, and agentic system behaviors**.
 - Hands-on automation scripting experience using **Python** or **JavaScript/TypeScript**.
 - Working familiarity with modern **CI/CD pipelines, Azure DevOps, and GitHub Enterprise**.
 - Familiarity with **testing frameworks, observability tools, or dashboarding software** (e.g., pytest, Playwright, LangSmith, PromptFlow, MLflow, Power BI, Grafana) is a major plus.
 - **Business-level Japanese (or above)** is mandatory for this role, as you must read Japanese business context documents and align directly with local teams.
-

Company Description