



AI Quality Scientist

Job Information

Recruiter

PROGRE Ltd

Hiring Company

AI Startup

Job ID

1560023

Industry

Internet, Web Services

Company Type

Small/Medium Company (300 employees or less)

Non-Japanese Ratio

About half Japanese

Job Type

Permanent Full-time

Location

Tokyo - 23 Wards, Shinjuku-ku

Salary

8 million yen ~ 16 million yen

Work Hours

10:00 ~ 19:00

Refreshed

May 13th, 2026 04:00

General Requirements

Minimum Experience Level

Over 6 years

Career Level

Mid Career

Minimum English Level

Business Level (Amount Used: English Only)

Minimum Japanese Level

None

Minimum Education Level

Post Grad Degree (PHD/MBA etc)

Visa Status

Permission to work in Japan required

Job Description

1. Evaluation Metric R&D

Research and implement LLM-as-Judge calibration (rubric design, bias detection, scoring rules).

Design and validate bespoke evaluation benchmarks to ensure construct validity.

Apply Reward Modeling and preference learning to improve evaluation accuracy.

2. Automated Pipeline Engineering

Build scalable automated evaluation pipelines integrated into CI/CD.

Develop agent evaluation harnesses supporting multi-turn dialogues, tool-use, and long-context scenarios.

3. Red Teaming & Safety

Automate adversarial testing and build policy compliance verification frameworks.

- Develop hallucination detection and calibration methods.

Required Skills

Minimum Qualifications

- Education: Master's degree or higher in CS, Machine Learning, Statistics, Physics, or related quantitative fields.
- Experience: 3+ years as an ML Engineer, Data Scientist, or Research Engineer.
- Technical: Proficiency in Python and ML frameworks (PyTorch, JAX, etc.).
- Domain Knowledge: Deep understanding of Generative AI evaluation (benchmarking, quantitative quality measurement).
- Language: Business-level English proficiency.

Preferred Qualifications

- Publication record at top-tier conferences (NeurIPS, ICML, ACL, etc.).
- Experience with RLHF, DPO, or preference learning.
- Expertise in AI Safety, Responsible AI, and automated red teaming.

Company Description