



## AI Quality Scientist

### 募集職種

人材紹介会社  
株式会社PROGRE

採用企業名  
AI Startup

求人ID  
1560023

業種  
インターネット・Webサービス

会社の種類  
中小企業 (従業員300名以下)

外国人の割合  
外国人 半数

雇用形態  
正社員

勤務地  
東京都 23区, 新宿区

給与  
800万円 ~ 1600万円

勤務時間  
10:00~19:00

更新日  
2026年05月13日 04:00

### 応募必要条件

職務経験  
6年以上

キャリアレベル  
中途経験者レベル

英語レベル  
ビジネス会話レベル (英語使用比率: 常時英語)

日本語レベル  
無し

最終学歴  
大学院卒 : 修士号/博士号

現在のビザ  
日本での就労許可が必要です

### 募集要項

#### 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.

---

## スキル・資格

### 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.

---

## 会社説明