

Michael Page

Research Engineer ・ Research Scientist (Responsible AI) (HAN)

日本語LLMの未来を支える、Responsible AI研究の最前線へ。

Job Information

Recruiter

Michael Page

Job ID

1603198

Industry

Bank, Trust Bank

Job Type

Permanent Full-time

Location

Tokyo - 23 Wards

Salary

6 million yen ~ 18 million yen

Refreshed

July 21st, 2026 13:13

General Requirements

Career Level

Entry Level

Minimum English Level

None

Minimum Japanese Level

None

Minimum Education Level

High-School

Visa Status

Permission to work in Japan required

Job Description

大規模言語モデル（LLM）の安全性・信頼性・セキュリティ向上を目的とした研究開発を担うポジションです。AIリスクの特定から対策技術の開発、社会実装、研究成果の論文文化までを推進し、安全な生成AIの実現を目指します。

Client Details

生成AIの研究開発および生成AIサービスの提供を行う企業です。技術パートナーシップの推進にも取り組み、特に日本語LLMの安全性・信頼性向上に注力した研究開発を進めています。2023年に設立された生成AI領域の専門企業です。

Description

- AI検査基盤（AI Red Teaming）の構築・運用・自動化
- SFT / RLHF / RLAIIFを用いた安全性学習手法の研究開発・最適化
- 有害表現検知や著作権侵害検知などGuardrails技術の開発
- 安全性関連データ基盤の設計・構築・運用
- AIリスクの特定およびリスク低減策の実装
- LLMの安全性・公平性・信頼性向上に向けた研究活動
- 研究成果の論文文化および知的財産化
- プロダクト開発チームや法務・コンプライアンス部門との連携

- 外部研究機関やパートナーとの共同研究・技術連携

Job Offer

- Responsible AIという注目領域の最先端研究に携われる
- 日本語LLMの安全性向上という社会的意義の高いテーマに取り組める
- AI Red TeamingやRLHF/RLAIFなど高度な研究開発経験を積める
- 研究成果を論文・特許として発信できる機会がある
- 法務・コンプライアンス部門と連携し、社会実装まで経験できる
- 年収650万円～1,800万円の幅広い報酬レンジ
- リモートワークと出社を組み合わせた柔軟な働き方
- フレックスタイム制度を活用可能
- 生成AI領域の研究者・エンジニアと協働できる環境

To apply online please click the 'Apply' button below. For a confidential discussion about this role please contact Keisuke Handa +81366276073.

Required Skills

- 計算機科学全般に関する深い知識
- 機械学習および大規模言語モデルに関する基礎知識・研究開発経験
- Responsible AI領域への強い関心と研究意欲
- 高度な専門内容を分かりやすく伝えるコミュニケーション能力
- 日本語ネイティブレベルおよび英語での論文読解・執筆能力
- 修士号（情報系または関連分野）または同等の実務経験

Company Description

事業内容：

生成AIの研究開発

生成AIサービスの提供

技術パートナーシップの推進

雇用形態：正社員

想定年収：650万円～1,800万円

勤務地：東京都港区海岸一丁目7番1号

働き方：リモート可（フルリモートは申請制／平均出社週1～2日）、フレックスタイム制あり