Remote AI Response Evaluator - Japanese

Be at the forefront of AI

Job Information

Hiring Company
Turing Enterprises Inc.

Job ID
1611530

Industry
Internet, Web Services

Company Type
International Company

Job Type
Contract

Location
Japan

Salary
Based on hourly rate

Refreshed
November 26th, 2026 11:41

General Requirements

Minimum Experience Level
Over 3 years

Career Level
Mid Career

Minimum English Level
Business Level

Minimum Japanese Level
Business Level

Minimum Education Level
Bachelor's Degree

Visa Status
Permission to work in Japan required

Job Description

Remote AI Response Evaluator - Japanese

How to apply:

Step 1:
Click the "Send Application/応募する" button on CareerCross

Step 2:
You will receive a confirmation email from CareerCross. Please use the external URL provided in the email to access the Turing website and complete your application.
***The same link can also be found at the bottom of this page.**

Role Overview:
As an **AI Quality Analyst**, you will evaluate a new personalization feature for Gemini. You will assess how well the model uses information from your past Gemini conversations, Gmail, Google Search, and YouTube activity to make responses more relevant and helpful.

This role requires a unique blend of creativity and analytical rigor. You will actively design prompts from the perspective of your own personal experiences.

You will then use your analytical skills to assess the quality of the model's personalized responses, evaluating dimensions like Grounding, Integration, and Helpfulness.

Description:

- In this role, you will be part of a dynamic team focused on evaluating the quality of personalized AI interactions. Your day-to-day work will involve:
- Designing and executing multi-turn conversational prompts (typically 1-5 turns) that require the AI to utilize your personal information and experiences.
- Evaluating model responses based on your intent from the starting prompt, checking if the personalization was appropriately applied.
- Analyzing responses for Grounding issues, ensuring claims about you are supported by evidence and not flawed inferences or hallucinations.
- Assessing Integration quality to ensure personal data is woven naturally into the response without robotic "overnarrating".
- Rigorously evaluating and stack-ranking two model responses side-by-side (SxS) to determine which is overall more helpful, easy to use, and enjoyable.
- Writing clear, defensible rationales for your comparisons, explicitly referencing where issues or positive aspects occurred in the conversation.
- Extracting and verifying "Debug Info" from the model to confirm that chat summaries and data sources were properly utilized.
- Maintaining strict data hygiene by deleting evaluation conversations to prevent them from polluting your future chat history.

Required Skills

Key Qualifications:

- Japanese Proficiency: Ability to read and write in Japanese with a high degree of comp, as Japanese is the focus language for this project.
- Personal Account Usage: Willingness to use your primary personal Google account (not a testing account) and enable personal data sources for a genuine assessment.
- Schedule Flexibility: Full-time availability in your local time zone is required. We are staffing a global, 24-hour operations team.

- Exceptional Analytical Thinking: Demonstrate ability to evaluate nuanced and ambiguous AI responses, specifically assessing personalization quality.
- Creative Prompt Engineering: Experience in designing creative, multi-turn starting prompts based on personal context to thoroughly test the model's capabilities.
- Strong Evaluation Acumen: Understanding of personalization concepts, including the ability to identify incorrect personalization, poor inferences, and forced connections.
- Meticulous Attention to Detail: The ability to review Side-by-Side (SxS) model responses and spot subtle differences in naturalness and overnarrating.
- Excellent Written Communication: Superior ability to write clear, concise, and structured rationales for model rankings, explicitly referencing specific turn numbers.
- Feedback: Ability to provide constructive feedback and detailed annotations.
- Communication: Excellent communication and collaboration skills.
- Independence: Self-motivated and able to work independently in a remote setting.
- Technical Setup: Desktop/Laptop set up with a good internet connection.

Education & Experience:

- BS/BA degree or equivalent experience in a relevant field (e.g., Policy, Law, Ethics, Linguistics, Journalism, Computer Science, or a related analytical field).
- Experience in data annotation, AI quality evaluation, content moderation, or a related role is strongly preferred.

Offer Details:

- **Commitments Required: at least 4 hours per day and up to 20 hours per week with 4 hours of overlap with PST.**
- Engagement type: Contractor
- Engagement Length: 3 months
- **Our offered rate for this project is $15 per hour.**

Evaluation Process:

- Shortlisted candidates will be sent a **Job Interest Form**.
- After the profile review, an **assessment** will be shared, which must be completed within **24 hours**.
- Based on the assessment outcomes, shortlisted candidates will be contacted to discuss the **pre-onboarding requirements**.

Step 2 link - Remote AI Response Evaluator - Japanese:
https://work.turing.com/invite/eyJhbGciOiJIUzI1NiIsInR5cCI6IkpXVCJ9.eyJqb2JJZCI6NjIxMzYsInVzZXJJZCI6OTkxLCJ1c2VyUm9sZSI6IlRBX1NPVVJDRUQiLCJpYXQiOjE3ODg3NzA1MDZ9.DNYuwCyL1bfzS2s9MXW7l3KeBYPbH2trzvkfAEfP
utm_source=job_board&utm_medium=+Career_Cross&utm_content=AI_Response_Evaluator&utm_campaign=Gemini_Personalization&utm_term=62136

Company Description