PREP Research Associate

This position is part of the National Institute of Standards and Technology (NIST) Professional Research Experience Program (PREP). NIST recognizes that its research staff may want to collaborate with researchers at academic institutions on specific projects of mutual interest and, therefore, requires those institutions to be recipients of a PREP award. The PREP program involves staff from a wide range of backgrounds conducting scientific research across various fields. Individuals in this position will perform technical work supporting the collaboration's scientific research.

Research Title:
Reliability of Human and LLM Annotations for AI Risk Assessment

The work will entail:
This project focuses on using Large Language Models (LLMs) to provide annotations of evaluation data (a.k.a., LLM as judge), and the design of an Inter-Annotator Agreement study to assess the reliability of both human and LLM annotations. The candidate will explore assessing the indicators of a given AI-related risk, determining how to identify them, and providing annotators with examples to annotate the presence of various risks. The project aims to develop an annotation framework for AI risk assessment and establish metrics for data quality in AI risk research, supporting broader work at NIST in assessing and measuring the validity and reliability of AI-related risks in data annotation.

U.S. Citizen Preferred

Key responsibilities will include but are not limited to:

- Gain familiarity with existing literature on data annotation and LLM as judge
- Understand NIST's role and ongoing efforts in assessing and measuring the validity and reliability of AI-related risks in data annotation
- Contribute to developing an annotation framework for AI risk assessment
- Collaborate effectively with cross-functional and interdisciplinary stakeholders to ensure successful project outcomes

Deliverables

- Contributions to a NIST report that supports ongoing NIST AI evaluation efforts focused on the design of an Inter-Annotator Agreement to assess the reliability of both human and LLM annotations.

<u>Qualifications</u>

- Background in Computer Science, Data Science, or related field.
- Education level: Bachelor's or Graduate Degree
- Strong interest in data annotation and AI risks
- Familiarity with scientific reading and technical writing

Privacy Act Statement

Authority: 15 U.S.C. § 278g-1(e)(1) and (e)(3) and 15 U.S.C. § 272(b) and (c)

Purpose: The National Institute for Standards and Technology (NIST) hosts the Professional Research Experience Program (PREP) which is designed to provide valuable laboratory experience and financial assistance to undergraduates, post-bachelor's degree holders, graduate students, master's degree holders, postdocs, and faculty.

PREP is a 5-year cooperative agreement between NIST laboratories and participating PREP Universities to establish a collaborative research relationship between NIST and U.S. institutions of higher education in the following disciplines including (but may not be limited to) biochemistry, biological sciences, chemistry, computer science, engineering, electronics, materials science, mathematics, nanoscale science, neutron science, physical science, physics, and statistics. This collection of information is needed to facilitate administrative functions of the PREP Program.

Routine Uses: NIST will use the information collected to perform the requisite reviews of the applications to determine eligibility, and to meet programmatic requirements. Disclosure of this information is also subject to all the published routine uses as identified in the Privacy Act System of Records Notices: NIST-1: NIST Associates.

Disclosure: Furnishing this information is voluntary. When you submit the form, you are indicating your voluntary consent for NIST to use of the information you submit for the purpose stated.