



About us

Humane Intelligence is a tech nonprofit building a community to professionalize the practice around algorithmic auditing and evaluation.

We are a programming platform environment for model evaluators and individuals seeking to learn more about model evaluations.

Mission

We are founded by industry veterans and optimists, Dr. Rumman Chowdhury and Jutta Williams.

We focus on safety, ethics, and subject-specific expertise and we aim to build the digital public infrastructure for model assessment.

A Better Way

We believe that every person has information, knowledge and lived experience to contribute.

Diverse issues with AI models are best solved by having more, diverse people contributing to solutions.

Contact Us to Learn More

info@humane-intelligence.org
www.humane-intelligence.org

Algorithmic Bias Bounty

The Challenge

- With the support of Google.org, we plan to conduct 10 "algorithmic bias bounty" programs over the next year.
- The program will center around a specific theme and participants will interact with the data via Humane Intelligence's Bias Bounty platform, with the purpose of red-teaming the model.

Goals

- Generate community norms for auditors/assessors.
- Provide ways of getting a broader community interested in the problem space.
- Help organizations solve problems.
- Create 'social proof' for auditors/assessors
- Help companies/regulators understand what skills to look for in an assessor.

Final Task

Create a probability estimation model that determines whether the prompt provided to a language model will elicit an outcome that demonstrates factuality, bias, or misdirection.