Description as given on AI Audit Challenge website https://hai.stanford.edu/policy/ai-audit-challenge

**What’s this about?**

* we lack the necessary tools to independently analyze and audit
* applied tools that can assess whether deployed AI systems exhibit bias or carry potential for discrimination.
* developing better tools for AI governance and in bridging the worlds of engineers and regulators, of technology and policy.

**What’s our objective?**

* to assess AI systems to determine whether they engage in prohibited discrimination.
* instead have chosen to prioritize impact through applied investigations, tools, and demonstrations.
* a particularly valuable area on which to concentrate is harmful bias in reference to protected categories
* How open source and deployed AI systems deal with protected characteristics and classes?
* Is it possible to identify indirect discrimination, through proxies and inferences?

Open-source modelssuch as GPT-NeoX-20B, BERT, GPT-J, YOLO, and PanGu-α.

Deployed systems in use by the public and private sector, such as COMPAS, GPT-3 and POL-INTEL.

* It is critical that policymakers and technology developers work in tandem.

**How is this project different from conferences like ACM FAccT?**

* In this case, our first target risk is bias and illegal discrimination.
* The outputs we are interested in are software, code, and/or tools that allow people to test publicly available algorithms and deployed models for illegal bias and discrimination, in ways that are useful and actionable for the people most likely to use such tools—namely, regulators, civil society, and journalists.

Evaluation criteria

* **Insights:** What did we learn using the tool?
* **Alignment:**How well anchored is the audit with legal and policy needs?
* **Impact:** How many people would benefit from the tool?
* **Ease of use:** Is the tool usable for our target audience?
* **Scalability:** Can the tool be used at scale and/or used in different contexts?
* **Replicability:** Can the results be replicated by other users using the same systems?
* **Documentation:** How well-explained are the findings?
* **Sustainability:** Is the tool financially and environmentally sustainable?