Ensuring Fairness with Transparent Auditing of Ethical Bias in AI Systems

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

2nd Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—With the rapid advancement of AI, there is a growing trend to integrate AI into decision-making processes. However, AI systems may exhibit biases that lead decision makers to draw unfair conclusions. Notably, the COMPAS system used in the American justice system to evaluate recidivism was found to favor racial majority groups; specifically, it violates a fairness standard called equalized odds. Various measures have been proposed to assess AI fairness. We present a framework for auditing AI fairness, involving third-party auditors and AI system provider, and we have created a tool to facilitate systematic examination of AI systems. Inclusion of independent and trusted auditors is pivotal for objectivity and accountability; third parties are often further necessary for their specialized expertise in relevant domains. Auditors, equipped with our tool, can thoroughly review AI systems for bias and fairness violations. The tool is open-sourced and publicly available. Unlike traditional AI systems, we advocate a transparent white-box and statistics-based approach. It can be utilized by third party auditors, AI developers, or the general public for reference when judging the fairness criterion of AI systems.**

***Index Terms*—keyword, keyword**

Identify applicable funding agency here. If none, delete this.