RESEARCH SUMMARY

Advancing Equitable Decisionmaking for the Department of Defense Through Fairness in Machine Learning

IRINEO CABREROS, JOSHUA SNOKE, OSONDE A. OSOBA, INEZ KHAN, MARC N. ELLIOTT

To access the full report, visit www.rand.org/t/RRA1542-1

ISSUE

Machine learning (ML) algorithms are increasingly used as an aid to human decisionmaking. However, there is a growing recognition that the use of ML algorithms may reinforce or exacerbate human biases, thereby perpetuating inequities. This situation is commonly referred to as *algorithmic bias*. The U.S. Department of Defense (DoD) is investing heavily in the development of ML algorithms to assist in many decisionmaking processes. At the same time, DoD has a strong stated interest in promoting diversity, equity, and inclusion (DE&I) at all levels of the organization. The goal of this report is to provide policymakers and developers of ML algorithms with a framework and tools to produce algorithms that are consistent with DoD's equity priorities. This report represents part of a larger effort to advance equity in DoD. Although predictive ML algorithms are deployed in some sectors within DoD—including intelligence and surveillance—ML algorithms are in the preliminary stages of development and are not at this time deployed in decisionmaking processes in the personnel space, where DoD has expressed equity goals. Despite this, we observe a growing interest in using ML algorithms as part of personnel decisions, as evidenced by the prototype tools developed in this space. Therefore, the utility of this report is primarily to preempt the possibility of algorithmic bias in eventual personnel decisionmaking applications within DoD rather than to address existing instances of algorithmic bias.

APPROACH

We provide a review of the written DoD policies and statements regarding DE&I in order to understand DoD's equity goals. We provide examples of the active development of ML technologies that interact with these equity goals, focusing specifically on ML algorithms that are embedded in decisionmaking processes. We review the technical concepts of algorithmic fairness and draw connections between DoD policy equity goals and possible comparable technical definitions of equity. We do not critique DoD policy statements as part of this work, though additional work could consider the adequacy of DoD's policies. We developed a framework and software tool, the RAND Algorithmic Equity Tool, to assist in the development of equitable predictive algorithms. For binary classification algorithms, this tool allows users to modify an algorithm to enforce specified equity goals. It also allows users to modify input training data to minimize the predictive influence of a protected attribute, such as race or sex. Importantly, the RAND Algorithmic Equity Tool helps users visualize trade-offs that are inherent to enforcing equity, such as diminished predictive accuracy. We display the functionality of this tool using a hypothetical ML algorithm that informs promotion decisions by automatically scoring candidates based on performance reviews, and we use the tool to enforce definitions of equity that may meet DoD's policy goals.

continued

CONCLUSIONS

With respect to DoD's equity goals, we identify three principles that may be linked to mathematical notions of equity: (1) career entry and progression should be free of discrimination with respect to protected attributes, including race, religion, or sex, (2) career placement and progression within DoD should be based on merit, and (3) DoD should represent the demographics of the country it serves. We argue that each of these principles corresponds to a notion of algorithmic fairness: specifically, *fairness through unawareness*, *true positive rate balance*, and *statistical parity*, respectively. To aid the development of equitable algorithms for particular decisionmaking processes, we propose the following five-stage procedure:

1. Determine equity risk.
2. Identify relevant equity mandates and priorities.
3. Determine relevant equity definitions.
4. Identify important performance priorities.
5. Weigh trade-offs of enforcing equity.

We show in our theoretical case study how this framework could be used to constrain algorithms to meet the identified DoD principles and the possible trade-offs with such constraints. In practice, it is not possible to satisfy each principle simultaneously, so priorities will need to be set. In our application of the RAND Algorithmic Equity Tool to a hypothetical case study, we show what it may look like to successfully enforce equity priorities.

RECOMMENDATIONS

Although ML algorithms have the potential to simplify existing human decisionmaking processes, there is a need to audit them to ensure that they do not result in inequitable outcomes. However, there is no universal approach to defining equitable outcomes; different decisionmaking processes involve different equity concerns. Additionally, attaining equity can come at the cost of other important priorities.

Therefore, the framework we recommend for developing equitable ML algorithms requires precisely defined equity and non-equity priorities. We emphasize that this required degree of precision is seldom available in the official mandates and statements provided by DoD regarding its DE&I priorities. To facilitate the development of equitable ML algorithms, DoD should collaborate with experts in the field of algorithmic fairness to translate institutional equity priorities into mathematical definitions. Once precise equity priorities are defined, the RAND Algorithmic Equity Tool allows users to enforce equity goals while monitoring the necessary trade-offs.

We propose that algorithms can and should be used as aids to human decisionmaking processes, both because algorithms can help reduce subjective human bias and because it is easier to audit and alter a well-constructed algorithm to enforce equitable outcomes. Although this report focuses on auditing algorithms, non-algorithmic (human only) processes can be similarly audited to ensure they are equitable. The findings in this report should be useful for framing the idea of equity, determining how to measure fairness, and collecting the right information in order to audit decisionmaking processes.

RAND PROJECT AIR FORCE

RAND Project AIR FORCE (PAF), a division of the RAND Corporation, is the Department of the Air Force's (DAF's) federally funded research and development center for studies and analyses, supporting both the United States Air Force and the United States Space Force. PAF provides DAF with independent analyses of policy alternatives affecting the development, employment, combat readiness, and support of current and future air, space, and cyber forces. For more information, visit PAF's website at www.rand.org/paf.