Jaiden Atterbury

Peter Gao

STAT 303

13 October 2022

Long Response 3: Forms of Resistance

The rapid development of algorithmic decision making systems and machine learning has led to powerful algorithms that have the potential to improve lives in sectors such as healthcare, where improved diagnostic speed and precision has allowed patients to be helped more efficiently. However, for more than a decade researchers have been warning the public about the dangers of relying on algorithms to make critical decisions on human lives. Ethical concerns such as biased outcomes, lack of transparency, and the misuse of personal data have led to the growing mistrust of these algorithms. It is well known in the wider research community that rather than remove bias, algorithms have codified and perpetuated it, while at the same time companies have continued to shield their algorithms from public scrutiny. This has led researchers to search for remedies and forms of resistance against these harms. One such form of resistance that tries to ensure that societal values are reflected in algorithms is the ever growing notion of algorithmic auditing.

Auditing in itself has been around for decades, as companies have long been required to issue audited financial statements for financial markets and other stakeholders. These audits were first put in place as companies’ internal operations were for the most part secretive. This gave an informational advantage over the public and other competing companies which could be abused through unethical misuse. Since algorithms lead to the same secrecy and unethical misuse of data as previously mentioned, lawmakers and researchers have advocated for algorithmic audits, which would similarly dissect and test algorithms to see how they work and whether or not they perform their stated goals without producing biased outcomes. In algorithmic auditing, there are two main actors: independent auditors who are hired from reputable firms offering algorithmic reviews, and algorithmic studies by researchers who weren’t hired by the company. Both of these actors attempt to provide reasonable assurance that the reports coming from the audited companies are free from misstatement. In general, audits proceed in a few different ways: by looking at an algorithm’s code and the data from its results, or by viewing an algorithm’s potential effects by examining each part of a model’s life cycle. This data is then used to assess whether the behavior is negatively impacting some interests/rights of people affected by that algorithm.

Since the idea of an algorithmic audit hasn’t been around for too long, there aren’t a plethora of examples about their use in practice, however there are a few historical examples of algorithmic auditing that exemplify their possibilities and limitations. The most famous example of a successful algorithmic audit comes from the MIT thesis by Joy Adowaa Buolamwini named *Gender Shades.* Gender shades, which has been coined as the first algorithmic audit of gender and skin type performance in commercial facial analysis models, was created to see how well different gender classification systems worked across different faces. In particular, it sought to see if the results of the classification changed based on somebody’s gender or skin type. The three companies that were evaluated in the study were: IBM, Face++, and Microsoft. Analysis of the results showed that companies had a high overall accuracy, but all companies performed better on males than females. They also performed better on lighter subjects than darker subjects. This study exemplifies the possibilities for change that an algorithmic audit can make, as within seven months of the original audit all three target companies released new versions of their models which reduced the accuracy disparities between males and females and darker and lighter skinned subgroups. While Gender Shades shines a light on the possibilities of algorithmic audits, the HireVue audit shows that there still exists plenty of limitations to the role algorithmic audits can play in mitigating ethical harms. In the audit, HireVue, a popular hiring software company, faced criticism that the algorithms it used to assess candidates through video interviews were biased. To combat this, HireVue called in the auditing firm O’Neil Risk Consulting & Algorithmic Auditing (ORCAA) to help assess their algorithm and find where bias may occur. HireVue said in a press release, that the audit found the software’s predictions “work as advertised with regard to fairness and bias issues.” Despite eliminating video from its interviews after the audit, HireVue was widely accused of using the audit as a PR stunt. The company implied that ORCAA’s audit fully cleared it of bias in its technology, however the audit in itself looked narrowly at a hiring test, and not HireVue’s candidate evaluation process as a whole. Articles published by Fast Company and MIT Technology Review called out the company for falsely representing the results of the audit.

It is important to note that algorithmic audits don’t have any restrictions on which ethical concerns can be addressed. Instead, an algorithmic audit can address whichever ethical concerns can be unveiled through tests/experiments run at each point in the model. The example of the aforementioned Gender Shades audit exemplifies how algorithmic audits can be effective at addressing ethical concerns presented by an algorithm. In fact, there are three primary ways in which algorithmic audits can be effective in addressing ethical concerns. The first of these possibilities created by algorithmic audits is the audits ability to set a corporation in the right direction in terms of model bias and data collection. Targeted algorithmic audits provide one mechanism to incentivize corporations to address the algorithmic bias present in data centered technologies that continue to have an ever growing role in daily life. Furthermore, because algorithmic audits encourage engagement with the issue of bias throughout the model-building process, they can facilitate a corporation's shift toward responsible data collection and use, while at the same time promoting better model development techniques. The next way in which algorithmic audits can effectively address ethical concerns is allowing users of audited algorithms to learn about how the algorithm works the way it does. Due to many companies’ algorithms being proprietary “black boxes,'' certain stakeholders may not understand how the algorithm came to a certain decision and therefore may not be confident in that decision’s fairness or accuracy. For example, university officials who used EAB’s Navigate advising software, which outputs a score that measures a student’s risk of not graduating on time, didn’t know how the scores were calculated nor how to use them. This sentiment is embodied by a quote from Carolyn Bassett, associate provost for student success at UMass Amherst who said “I certainly haven’t had a lot of information from behind the proprietary algorithms.” However, due to the algorithmic audits nature of unveiling systematic biases and deviations from expected output throughout every stage of the model, algorithmic audits are very effective at explaining to users of an algorithm why and how the algorithms do what they do, and if they should or shouldn’t trust the decision made by the algorithm. The last, and some argue the most important way algorithmic audits can effectively mitigate ethical harms created by an algorithm is the transparency and accountability they provide. Independent research audits, like the one done by ProPublica on COMPAS, allow full disclosure of results found in the tests/experiments run on the algorithm without intervention from the private agency who developed the algorithm. This allows more transparency and awareness about the problems presented by the algorithm to the general public. In turn, this awareness can lead to other forms of resistance such as consumer boycotts and worker walkouts.

Although auditing algorithms is the industry standard for addressing ethical concerns present in the algorithms used by many companies big and small, there are still limitations to what algorithmic audits can do and how effective they can be. In particular there are three overarching limitations to algorithmic auditing that need to be addressed in order to fully evaluate their effectiveness. First off, algorithmic audits are usually very limited in scope. Despite the fact that audits on algorithms can address any ethical concerns unveiled in the tests run on the algorithms, current ethical assessments of algorithms focus on very specific and technical notions of fairness/transparency that do not consider a broader social context. This is due to the fact that auditors can’t test everything, thus audits usually look at only a single system without including how it interacts with other systems present in the current domain. The next limiting factor that keeps algorithmic audits from running at their peak potential is the fact that auditors don’t always get access to all of the information surrounding a particular algorithm. Since tech companies don’t want to reveal the proprietary information/formulas in their technology that sets them apart from their competitors, even with the auditors they hire, these auditors are put in scenarios where they don’t have access to the model's code. Without access to this proprietary information, auditors can’t create finished evaluations nor can they validate any specific part of an algorithm of being free from bias. Furthermore, auditors who don’t have access to proprietary information risk breaching company Terms of Service as well as face uncertainty around how corporations will react to their findings. Due to these risks, much algorithmic audit work has steered away from directly challenging companies to change the way their systems work and focus more on public awareness of any found problems. The last way in which there exists limitations in the effectiveness of algorithmic auditing is the lack of regulation surrounding auditing itself. Hiring an auditor isn’t common practice among companies who use algorithms as companies have no obligation to do so. Furthermore, companies don’t want to deal with any backlash from results that may come from audits. For those that do hire auditors, there are no industry standards for what an audit should entail. For example, in January 2018, New York City became the first U.S. jurisdiction to enact a law creating a task force to provide recommendations on the use of algorithmic decision systems, however the law doesn’t spell out how the audits should be formed or conducted. With this said, even with a task force that maps out what algorithmic audits should entail, there are no industry standards/regulations that hold the companies to make any changes suggested by the audits.

With the rapid presence of machine-learning and algorithmic decision making systems, it is ever so important that we take a step back and truly analyze the ethical concerns these technologies may produce. Due to these widespread ethical concerns, algorithmic auditing acts as a vessel through which we can combat these same concerns. It is true that there are many limitations to what algorithmic audits can achieve, as most of the time results can only provide a glimpse of a much bigger picture. While algorithmic audits provide a level of necessary oversight in the growing field of algorithmic decision making systems, they are not a “one size fits all” solution to the problems these technologies can yield. However, it is important to note that some oversight is better than none and algorithmic auditing provides a direct way for researchers to find systematic bias in models and push for change, something that consumer boycotts and worker walkouts can’t achieve by themselves.

Works Cited

Angwin, Julia, et al. “Machine Bias.” *ProPublica*, 23 May 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=jRdnwabwdw5HLiHY-R3nqWS5DOjEM7W->.

Baeza-Yates, Ricardo. “Wielding The Double-Edged Sword Of Algorithmic Auditing.” *Forbes*, Forbes Magazine, 25 July 2022, <https://www.forbes.com/sites/forbestechcouncil/2022/07/25/wielding-the-double-edged-sword-of-algorithmic-auditing/?sh=3343aca330c9>.

Brown, Shea, et al. “The Algorithm Audit: Scoring the Algorithms That Score Us.” *SAGE Journals*, Sara Miller McCune, 28 Jan. 2021, <https://journals.sagepub.com/doi/full/10.1177/2053951720983865>.

Buolamwini, Joy, and Inioluwa Deborah Raji. “Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products.” *MIT Media Lab*, 24 Jan. 2019, <https://www.media.mit.edu/publications/actionable-auditing-investigating-the-impact-of-publicly-naming-biased-performance-results-of-commercial-ai-products/>.

Buolamwini, Joy. “Project Overview: Gender Shades.” *MIT Media Lab*, Feb. 2018, <https://www.media.mit.edu/projects/gender-shades/overview/>.

Engler, Alex C. “Outside Auditors Are Struggling to Hold AI Companies Accountable.” *Fast Company*, Fast Company, Inc, 26 Jan. 2021, <https://www.fastcompany.com/90597594/ai-algorithm-auditing-hirevue>.

Feathers, Todd. “Major Universities Are Using Race as a ‘High Impact Predictor’ of Student Success.” *The Markup*, 2 Mar. 2021, <https://themarkup.org/machine-learning/2021/03/02/major-universities-are-using-race-as-a-high-impact-predictor-of-student-success>.

Guszcza, et al. “Why We Need to Audit Algorithms.” *Harvard Business Review*, Harvard Business Publishing, 28 Nov. 2018, <https://hbr.org/2018/11/why-we-need-to-audit-algorithms>.

Kassir, Sara. “Algorithmic Auditing: The Key to Making Machine Learning in the Public Interest.” *IBM Center for The Business of Government*, Dec. 2019, <https://www.businessofgovernment.org/sites/default/files/Algorithmic%20Auditing.pdf>.

MIT Media Lab. “Gender Shades.” *YouTube*, YouTube, 9 Feb. 2018, <https://www.youtube.com/watch?v=TWWsW1w-BVo>.

Ng, Alfred. “Can Auditing Eliminate Bias from Algorithms?” *The Markup*, 23 Feb. 2021, <https://themarkup.org/the-breakdown/2021/02/23/can-auditing-eliminate-bias-from-algorithms>.

Richardson, Rashida. “History of the Legislative Process.” *Confronting Black Boxes:A Shadow Report of the New York City Automated Decision System Task Force*, New York, NY, 2019, pp. 11–12.

Sarmah-Hightower, Satta. “3 Reasons Your Organization Will Need External Algorithm Assessors.” *Forbes*, Forbes Magazine, 26 Oct. 2022, <https://www.forbes.com/sites/kpmg/2022/10/26/3-reasons-your-organization-will-need-external-algorithm-assessors/?sh=4bf98c6ab517>.

Schellmann, Hilke. “Auditors Are Testing Hiring Algorithms for Bias, but There’s No Easy Fix.” *MIT Technology Review*, MIT Technology Review, 11 Mar. 2021, <https://www.technologyreview.com/2021/02/11/1017955/auditors-testing-ai-hiring-algorithms-bias-big-questions-remain/>