From AI Audit to Accountability: Understanding the Policy Perspectives Required for Accountability

Abstract

Some of the most horrific anecdotes about algorithmic harms and injustice are tired to pre-mature AI deployments leading to social and benefits injustice. The AI audit ecosystem remains fragmented with tools and frameworks scattered and, in some cases closed and gatekept. Giving the increasingly visible policy developments mandating audits and the proliferation of algorithmic products, algorithmic audits are increasingly critical tools for holding vendors and operators accountable. But without an understanding of the drivers of AI audit relating to a particular sector, it is hard for meaningful policy to drive and hold operators accountable.

We propose to map the overall AI audit ecosystem trends, and their relationship with societal and technological innovation and identify policy mechanisms and framework and practical recommendations for policy developments and future research to advance the AI Audits for accountability.

Background

We have seen reports of blockage and access to health [1], housing [2] unemployment fraud algorithms [3] [4] and law enforcement [5]. There have been calls for the need for "technological due process" — the importance of both understanding what's happening in automated systems and being given meaningful ways to challenge them[1]. Also, critics have pointed out that, when designing these programs, incentives are not always aligned with easy interfaces and intelligible processes. Audits are evaluations with an expectation for accountability [6]. With audits, it's not enough to just make a reliable measurement about how well a system is working, but also to ask, have this measurement lead to consequential outcomes that can further protect those impacted from any potential harm [8]. Evaluation is the process of assessment and measurement, to vet either explicit or implicit claims of performance or compliance to standards. Accountability, on the other hand, involves the informed and consequential judgment of the actions of decision-making actors within a system [7].

1

Although most audits share the goals of measuring and translating meaningful assessments into some form of accountability, the evaluations themselves can take on multiple forms, given the constraints and objectives of the various stakeholders that participate in audit activities.

AI algorithms causing real-world consequences, such as unfairly denying someone medical care or unjustly advising against parole. Also, researcher access to platform data is important. While in some instances researchers have failed to separate the training data to test the test's performance. This results in data leakage and a disaster in itself.

Already in the US conversations by government agencies regarding algorithms is already taking shape towards the congress. The federal Trade Commission in their report proposes legislation focused on 'the transparency and accountability of platforms' to be known as the platform accountability and transparency act [9].

The humans in the loop approach have been used also to supervise algorithms but this is insufficient in itself. There needs to be a framework to audit the system for accountability. Researchers, academics and policymakers have long implored platform companies to be more transparent, including around the core algorithms for example, those used for select and personalizing content. In the absence of a unified framework for auditing that also emphasizes accountability, legislative proposals have emphasized on algorithmic transparency and accountability. Yet, it's not obvious what transparency actually means for such large and complex systems where outcomes are the result of interactions between algorithms and people. Open-sourcing code could be useful but AI systems are complex, and there isn't a single algorithm that underpins the whole system. Rather, different machine learning models trained on different data sets perform a variety of functions within the system.

Our proposition is that a policy discussion, formulation and regulation for AI Audit and accountability should be driven by the fundamental understanding on transparent systems especially for *documentation, code, data and research*.

The AI audit ecosystem remains fragmented with tools and frameworks scattered and, in some cases closed and gatekept. Giving the increasingly visible policy developments mandating audits and the proliferation of algorithmic products, algorithmic audits are increasingly critical tools for holding vendors

and operators accountable. But without an understanding of the drivers of AI audit relating to a particular sector, it is hard for meaningful policy to drive and hold operators accountable.

Our research hopes to address the need for AI accountability policy instruments to advance the AI Audits. The project aims to:

1. Identify overall AI audit ecosystem trends, and their relationship with societal and technological innovation.

2. Identify policy mechanisms and framework and practical ways in which policy developments and research need to think about policy documents to support AI audit outcomes.

3. Propose practical and tangible questions and recommendations for those working on AI legislation and internal/external audit stakeholders and set the scene for future policy research.

Goals and objectives

The goal is to support and facilitate the design and development of policy instruments, mechanisms, programs leading impactful legislation for accountable AI systems. To map the AI Audit ecosystem, stakeholders and macro-level trends combined with accountability challenges. And describe the strategies and roadmap for accountability.

Research questions

The central research question is: How might we codify the policy perspectives required for AI accountability?

Preliminary questions are; a) Mapping the AI Audit ecosystem, identify policy frameworks and mechanisms (b) how they interact with societal and technological advancements, and c) how this intersection could impact AI regulation and governance for accountability.

Methodology

Phase 1: In-depth systematic literature review and analysis to document and map the global AI audit ecosystem trends, review and analysis of policy mechanisms and frameworks to address accountability and transparency and impact of accountability frameworks in AI auditing on regulations and governance.

Phase 2: Resulting policy draft will be subjected to workshop and peer review sessions and areas and sectors of focus identified for further analysis.

Phase 3: key informant interviews and working group discussions

Phase 4: Finished draft policy document and published collection of paper

Scope of work

We'll publish an open-source compendium of collected short papers will focus primarily on African researchers and contributors. Special outreach and solicitation will be sought from women and minority communities.

Collection will focus on mainly on public sector AI and Algorithmic Decision-making systems and involved various sectors, stakeholders and contexts. Some of the broad topics and themes to be considered in the collection include:

- Interpretation and leveraging audit results for accountability
- The politics of AI Audit
- Policy recommendations for AI Audits in Africa
- Strategies for engaging policymakers and stakeholders in AI Audit and Accountability
- Understanding the African policy ecosystem for effective AI Auditing
- Reproducibility & Accountability crisis in AI
- Reproducibility & Accountability crisis in AI: African perspectives on AI Auditing

Endnotes:

[1] https://www.theverge.com/2018/3/21/17144260/healthcare-medicaid-algorithm-arkansas-cerebral-palsy

[2] https://themarkup.org/locked-out/2020/05/28/access-denied-faulty-automated-background-checks-freeze-out-renters

[3] https://undark.org/2020/06/01/michigan-unemployment-fraud-algorithm/

[4] Charette, Robert. "Michigan's MiDAS Unemployment System: Algorithm Alchemy Created Lead, Not Gold-IEEE Spectrum." *IEEE Spectrum* 18.3 (2018): 6.

[5] Hill, Kashmir. "Wrongfully accused by an algorithm." *The New York Times* 24 (2020).

[6] Raji, Inioluwa Deborah. "The Anatomy of an AI Audit." 2022 (Forthcoming).

[7] Raji, Inioluwa Deborah, and Joy Buolamwini. "Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products." Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. 2019.

[8] Wieringa, Maranke. "What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability." *Proceedings of the 2020 conference on fairness, accountability, and transparency*. 2020.

[9] https://techpolicy.press/document-senators-propose-platform-accountability-and-transparency-act/