Who audits the Auditors? Challenges and paths for auditing Artificial Intelligence

The growing adoption of Artificial Intelligence (AI) systems in government and private sectors has intensified concerns about bias, transparency, and the ethical impact of these technologies. To mitigate these risks, algorithmic auditing practices have been implemented to ensure the compliance and fairness of AI systems. However, a crucial question arises: who audits the auditors of these systems themselves? This study investigates the algorithmic audit ecosystem to understand the oversight and accountability mechanisms applied to AI auditors. The aim is to identify gaps and propose recommendations to improve the effectiveness and reliability of algorithmic audits, considering the current regulatory landscape, mainly in the European Union, whose guidelines can serve as a reference for other jurisdictions.

The research addresses the importance of audits in the government sector, highlighting the need for specific structures to audit machine-learning projects in public agencies. Outsourcing these services to private companies raises ethical and technical concerns, as the lack of centralized oversight can amplify social inequalities. The approach is complemented by an analysis of the performance of private companies, highlighting that biased and flawed systems can negatively impact marginalized communities.

In the regulatory context, the European Union's Artificial Intelligence Act stands out. This act establishes a comprehensive legal framework for AI, classifying systems according to levels of risk and imposing stricter requirements for high-risk ones, such as security components in critical infrastructures and recruitment tools. However, the effectiveness of these regulations depends on the ability to audit not only AI systems but also the auditors themselves, who are responsible for these assessments.

One central challenge is ensuring that algorithmic audits do not reproduce the same biases and flaws they seek to correct. Many studies point out that AI auditing is still fragmented, with no standardization and little transparency about who evaluates the auditors and their methods. One of the main problems identified is the lack of accountability and transparent mechanisms for holding auditors to account, especially when they involve decisions that affect the lives of millions of people.

The algorithmic auditing literature highlights several key aspects for effectiveness. Läm et al. (2024) emphasize the need to expand the audit scope beyond algorithm accuracy to include social impacts, particularly in areas like targeted advertising. They advocate for iterative and collaborative audit models, integrating human auditors and automated systems to minimize risks. Additionally, involving independent researchers and civil society is suggested to enhance transparency and fairness in audits.

Another relevant issue the research addresses is the continuous auditing of AI systems in operation. The concept of "dynamic auditing" has been advocated as a necessary mechanism to prevent AI systems from becoming obsolete or acquiring new biases after implementation. Post-market monitoring, as provided for in the European guidelines, is a step in this direction. Still, its effectiveness will depend on enforcement capacity and mechanisms to correct identified flaws.

The main results of the research indicate that the absence of supervision over auditors can compromise the integrity of audits, allowing biases and inadequate practices to go

undetected. In addition, it was observed that many audits are financed by the AI companies themselves, raising doubts about their impartiality. The conclusion is that it is imperative to develop auditing frameworks that include mechanisms for oversight and accountability of AI auditors, ensuring that evaluations are conducted ethically, transparently, and fairly.

Finally, European regulatory guidelines, especially the Artificial Intelligence Act, serve as relevant case studies to assess the feasibility of regulations that ensure effective audits. It is argued that AI auditing should be seen not only as a technical mechanism but also as an ethical and political imperative, fundamental to ensuring that Artificial Intelligence benefits society in an equitable and responsible manner.

References

BANDY, Jack. *Problematic machine behavior: a systematic literature review of algorithm audits*. Proceedings of the ACM on Human-Computer Interaction, v. 5, n. CSCW1, p. 1–34, 2021. Available at: https://doi.org/10.1145/3449148. Accessed on: Mar. 1, 2025.

DEY, Diptish; BHAUMIK, Debarati. *APPRAISE: a governance framework for innovation with AI systems*. Computers and Society (cs.CY), 2023. Available at: https://arxiv.org/abs/2309.14876. Accessed on: Mar. 1, 2025.

FERGUSSON, Grant. *Outsourced automated: how AI companies have taken over government decision-making*. Electronic Privacy Information Center. Available at: https://epic.org/wp-content/uploads/2023/09/FINAL-EPIC-Outsourced-Automated-Report-w-Appendix-Updated-9.26.23.pdf. Accessed on: Mar. 1, 2025.

GANAPINI, Marianna; PANAI, Enrico. *An audit framework for adopting AI nudging on children*. The Ethics Lab, 2023. Available at: https://arxiv.org/pdf/2304.14338. Accessed on: Mar. 1, 2025.

KENNEDY, Alayna. *Why we need to audit government AI*. Montreal AI Ethics Institute, 2024. Available at: https://montrealethics.ai/why-we-need-to-audit-government-ai/. Accessed on: Mar. 1, 2025.

LÄM, Michelle S. et al. *Sociotechnical audits: broadening the algorithm auditing lens to investigate targeted advertising*. 2024. Available at: https://dl.acm.org/doi/10.1145/3610209. Accessed on: Mar. 1, 2025.

MAUGHAN, Krystal; NGONG, Ivoline C.; NEAR, Joseph P. *Prediction sensitivity: continual audit of counterfactual fairness in deployed classifiers*. 2022. Available at: https://arxiv.org/pdf/2202.04504. Accessed on: Mar. 1, 2025.

MÖKANDER, J.; AXENTE, M.; CASOLARI, F. et al. *Conformity assessments and post-market monitoring: a guide to the role of auditing in the proposed European AI regulation*. Minds & Machines, v. 32, p. 241–268, 2022. Available at: https://doi.org/10.1007/s11023-021-09577-4. Accessed on: Mar. 1, 2025.

RASTOGI, Charvi et al. *Supporting human-AI collaboration in auditing LLMs with LLMs*. 2023. Available at: https://arxiv.org/pdf/2304.09991. Accessed on: Mar. 1, 2025.

REGULATION (EU) 2024/1689 of the European Parliament and of the Council, June 13, 2024. *EUR-Lex*, 2024. Available at: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689. Accessed on: Mar. 1, 2025.

VECCHIONE, Briana; LEVY, Karen; BAROCAS, Solon. *Algorithmic auditing and social justice: lessons from the history of audit studies*. In: EAAMO '21: Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization. New York: ACM, 2021. p. 1–9. Available at: https://doi.org/10.1145/3465416.348329. Accessed on: Mar. 1, 2025.

WÖRSDÖRFER, M. *The E.U.'s artificial intelligence act: an ordoliberal assessment*. AI Ethics, v. 5, p. 263–278, 2025. Available at: https://doi.org/10.1007/s43681-023-00337-x. Accessed on: Mar. 1, 2025.

WRIGHT, Connor. *Who audits the auditors? Recommendations from a field scan of the algorithmic auditing ecosystem*. Montreal AI Ethics Institute, 2024. Available at: https://montrealethics.ai/who-audits-the-auditors-recommendations-from-a-field-scan-of-the-algorithmic-auditing-ecosystem/. Accessed on: Mar. 1, 2025.