Ada Lovelace Institute

Discussion paper

Code & conduct

How to create third party auditing regimes for AI

Emerging technology & industry practice

June 2024

Contents

Executive summary

People expect artificial intelligence (AI) systems that impact their lives to be safe, effective and compliant with local and national laws.[1][2] But currently, most AI systems are not required to undergo independent testing for the range of risks they can pose to the public. As these technologies become more widely integrated into our daily lives, the current rate of adoption of AI is vastly outpacing the ability of developers, regulators or civil society organisations to ensure they are safe.

For AI to benefit people and society, it is essential that these technologies are tested iteratively pre- and post-deployment to ensure they are safe, effective and legal. In other safety-critical domains – like medicines, transport and automotive safety – auditing and assessment have helped demonstrate that products are safe to use and that defective products can be recalled from the market. A robust assurance-services market has grown as these domains have matured. This has been crucial for fostering public trust in the institutions building and operating these products and technologies, ensuring a high standard of safety and making sure that the efficacy of these systems is well understood.

AI systems should be held to this same standard. The challenge facing policymakers is how to construct a regulatory and governance ecosystem to assess and mitigate the range of risks these systems may raise for people and society.

One proposed method is the use of **algorithm audits**, which involve testing an AI product or service for certain kinds of risks – such as whether the system discriminates against certain demographic groups or produces toxic or offensive content.

1 Ada Lovelace Institute, *What do the public think about AI?* (2023), Available at: https://www.adalovelaceinstitute.org/evidence-review/what-do-the-public-think-about-ai/ (Accessed: 17 May 2024);

2 Ada Lovelace Institute and The Alan Turing Institute, *How do people feel about AI?* (2023), Available at: https://www.adalovelaceinstitute.org/report/public-attitudes-ai/ (Accessed: 17 May 2024);

What is an algorithm audit?

As a practice, algorithm audits stem from academic and civil society research, which determines several important points about the use of algorithm audits.

- **Algorithm audits are a necessary, but not sufficient, method for enabling accountability.** Conducting an audit can provide regulators, developers and users of an AI system with some degree of assurance of an AI system's performance. However, developers do not necessarily take action – based on the results of an audit – to reduce harm, remove products from the market, or provide redress for affected parties.[3]

- **Auditing may have to occur routinely throughout the lifecycle of AI systems.** Many risks from AI systems can originate or proliferate at different stages of a system's lifecycle, from the data collection stage all the way through to how human operators use or misuse a system. For example, an AI system tested in 'lab settings' may appear not to lead to discriminatory outcomes, but it may be used in a way that causes this outcome when deployed in specific contexts.[4] It is necessary to routinely if not continuously audit an AI system's behaviour to understand its impacts.

- **Conducting an algorithm audit requires an auditor to have a clear remit and a standardised test for assessment.** For some risks like algorithmic bias, there is a lack of standardised tests to identify bias in different contexts.[5]

- **Audits should involve more than just a technical evaluation of the system itself and should also involve inspection or assessment of organisational processes and policies.**[6] Their scope should also include reviewing how the introduction of that system may impact the environment in which it is used (such as how it changes the behaviour of human operators).

3 Birhane et al., (2024), 'AI auditing: The Broken Bus on the Road to AI Accountability', arXiv, Available at: http://arxiv.org/abs/2401.14462 (Accessed: 31 January 2024);

4 Ada Lovelace Institute, *Safe before sale*, (2023), Available at: https://www.adalovelaceinstitute.org/report/safe-before-sale/ (Accessed: 17 May 2024);

5 Ada Lovelace Institute, *AI assurance? Assessing and mitigating risks across the AI lifecycle*, (2023), Available at: https://www.adalovelaceinstitute.org/report/risks-ai-systems/ (Accessed: 11 October 2023);

6 Mökander and Floridi, (2022), 'Operationalising AI governance through ethics-based auditing: an industry case study', doi: 10.1007/s43681-022-00171-7;

- **Audits conducted by independent auditors tend to result in higher quality evaluations than assessments conducted by internal teams.** However, independent auditors face serious challenges around access to the data and information that is necessary to complete an audit, and there is no consensus on how to define an 'independent auditor'.[7] There is a long history of audits acting as a form of regulatory arbitrage, which must be addressed through the introduction of safeguards, standards and oversight.[8]

- **Audits are just one tool in a toolbox that can complement other accountability methods.** If used well, audits can complement and enhance other AI accountability and governance methods like algorithmic impact assessments, human rights impact assessments and regulatory inspections.[9]

Algorithm audits in practice

Algorithm auditing requirements are beginning to appear in local, national and multinational AI regulations. Some national regulators including the Australian Competition and Consumer Commission[10] and the Netherlands Court of Audit[11] have the capability to audit certain algorithms to assess compliance with local laws, but these have been used sparingly.

Auditing powers for regulators have also been specified in emerging online safety legislation, including powers for Ofcom under the UK's Online Safety Act. The European Union's Digital Services Act specifies requirements for a third-party auditing regime where 'Very Large Online Platforms' must undergo regular audits by an independent auditor to ensure the platform is applying its content moderation policies. Independent algorithm auditing has also been proposed by the UK Government's Responsible Technology Adoption Unit as one part of its AI assurance framework.[12]

---

7 Raji et al., (2022), 'Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance', arXiv, Available at: http://arxiv.org/abs/2206.04737 (Accessed: 8 August 2022);

8 Terzis, Veale and Gaumann, (2024), 'Law and the Emerging Political Economy of Algorithmic Audits', doi: 10.31219/osf.io/xvqz7;

9 Ada Lovelace Institute, (2022), Algorithmic impact assessment: a case study in healthcare, Available at: https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/ (Accessed: 13 June 2023);

10 Commission, (2020), Trivago misled consumers about hotel room rates, Available at: https://www.accc.gov.au/media-release/trivago-misled-consumers-about-hotel-room-rates (Accessed: 17 May 2024);

11 Rekenkamer, (2022), An Audit of 9 Algorithms used by the Dutch Government - Report - Netherlands Court of Audit, Algemene Rekenkamer, Available at: https://english.rekenkamer.nl/publications/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government (Accessed: 18 January 2024);

12 Portfolio of AI assurance techniques - GOV.UK, (no date), Available at: https://www.gov.uk/guidance/portfolio-of-ai-assurance-techniques (Accessed: 17 May 2024);

This project studied the experience of auditors conducting independent bias audits of companies using AEDTs

Introducing our research

As national governments seek to integrate independent algorithm auditing regimes into proposals for regulating AI systems, it is crucial they learn the lessons from the 'first wave' of these laws. In late 2023, the Ada Lovelace Institute and Data & Society conducted qualitative research into the first attempt to create an independent algorithm auditing regime for commercial machine learning systems.

This law – New York City's Local Law 144 (LL 144) – requires employers in the city who use automated employment decision-making tools (AEDTs) to commission an independent bias audit and make the results publicly available. AEDTs are defined as tools that 'substantially assist' in the hiring of job candidates, such as tools that automatically sift job applicant CVs or analyse personality traits from interview video. The audits focus on identifying potentially biased outcomes based on race and gender, using a measure known as 'disparate impact' that compares rates of hiring across demographic groups.

This project studied the experience of auditors conducting independent bias audits of companies using AEDTs. We conducted 15 interviews with 17 practitioners and experts to explore the following research questions:

- **RQ1:** What are the practical components of a bias audit in this context?
- **RQ2:** What are the components, relationships and incentives that make for an effective bias auditing regime?
- **RQ3:** What are the experiences of auditors, and how can we use those experiences to inform wider policy and practice around other algorithm auditing regimes?

Findings

Our research surfaced several important findings for policymakers who are designing future independent algorithm auditing regimes:

- LL 144 resoundingly failed to create a robust third-party auditing ecosystem that improved fairness outcomes in hiring. In total, we could identify only around 20 employers in all of New York City that completed and published an AEDT audit under this law.

- Some of the challenges related to the law itself, including a faulty 'theory of change' that did not require companies to stop the use of biased tools but only to make the results of the audit available for job candidates.

- The law also suffers from a narrow and gameable definition of what systems were in scope and a lack of meaningful mechanisms for enforcement by regulators.

- Other reasons for the law's apparent failure in some key areas came down to the complex dynamics around algorithm auditing. These included cultural and practical challenges that auditors faced to get the data necessary to conduct audits, and a lack of clear standards for what roles and practices auditors should adopt.

- On the positive side, LL 144 successfully created a standardised audit test for auditors to employ. According to auditors, its introduction caused some companies to adopt wider responsible AI practices that they might not otherwise have adopted.

Recommendations

Following our findings, we offer **six recommendations** for policymakers designing future algorithm auditing regimes:

Recommendation 1: Auditing laws must establish clear definitions that clearly capture the full range of AI systems in scope

These should include defining which AI systems should be audited, how an audit should be conducted and by whom. We propose that members of affected communities and civil society organisations should be closely consulted when creating these standards, metrics and definitions.

Recommendation 2: Auditing laws must establish clear standards of practice on the role and responsibilities of auditors

Beyond setting out the requirements for and components of the audit, policymakers should also create standards establishing appropriate auditing practice, the specific role of an auditor and mechanisms for auditor oversight.

Recommendation 3: Auditing laws must enable smooth data collection for auditors

Auditors routinely cite data access issues as the greatest obstacle for conducting audits. Auditing laws must adopt clear procedures and requirements around data access to conduct the relevant tests, including information and documentation about the datasets that are turned over.

Recommendation 4: Auditing laws must establish meaningful metrics that accurately capture a risk

Policymakers must develop region-specific and risk-specific audit metrics, acknowledging that not all risks can be quantified into a metric.

Recommendation 5: Audits should follow a theory of change that results in meaningful outcomes

The design of the audit should facilitate meaningful outcomes for people and society. Audits should be publicly accessible and legible for lay audiences via a transparency register, along with mechanisms that enable people to query or challenge an audit result, in order to create accountability.

Recommendation 6: Auditing laws need mechanisms to monitor and enforce against non-compliance.

Sanctions for non-compliance must be serious and substantial. Companies may view weaker penalties as a 'cost of doing business' or may choose to wait to see if a regulator begins to monitor and enforce penalties before taking action. Regulators will require powers to inspect companies suspected of failing to comply with independent auditing regimes, including powers to require disclosure of documents or details about their use of AI systems. Regulators must be sufficiently resourced to hire staff to conduct enforcement operations.

We urge policymakers to adopt independent auditing regimes for AI systems and hope these recommendations will help ensure future regimes create safer and more effective AI systems.

Algorithm auditing remains uncommon and is used in an ad-hoc manner for some kinds of AI systems

Introduction

In sectors like food safety, drug development and aviation, independent auditing regimes have helped ensure that systems and services are safe for the public to use.[13] Like these sectors, AI is another domain of high-risk and high-stakes decision making. It is important that policymakers build mechanisms that assure people that AI systems are safe, legal and effective.[14]

One tool in the toolbox for ensuring the safety, efficacy and legality of AI systems is the use of algorithm audits. Algorithm audits primarily stem from the academic literature around human–computer interaction and fairness, accountability and transparency in machine learning.[15] [16] In these studies, researchers and civil society organisations would assess an AI system's behaviour for issues like bias or toxicity, usually without any direct access to the underlying system.[17] [18]

Some teams within major technology companies have adopted algorithm auditing practices to assess their systems for various risks. There is also a nascent industry of second-party auditors who offer this service for hire.[19] A second-party audit is where the audit is commissioned by a company but conducted by a separate organisation. For example Meta commissioned a civil rights audit of its platform in 2020, which found failures to protect civil rights on a range of issues, from voter suppression to hate speech.[20]

13 Raji et al., (2022), 'Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance', arXiv, Available at: http://arxiv.org/abs/2206.04737 (Accessed: 8 August 2022);

14 Ada Lovelace Institute, 'Mission critical', (2023), Available at: https://www.adalovelaceinstitute.org/policy-briefing/ai-safety/ (Accessed: 1 November 2023);

15 Buolamwini and Gebru, (2018), 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', Conference on Fairness, Accountability and Transparency, PMLR, Available at: https://proceedings.mlr.press/v81/buolamwini18a.html (Accessed: 17 May 2024);

16 Sandvig et al., (no date), 'Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms';

17 Angwin et al., (2016), Machine Bias, Available at: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing?token=J5k3utYNqmWvBjTaTBs4TylpiUAiFx2o (Accessed: 31 March 2021);

18 Obermeyer et al., (2019), 'Dissecting racial bias in an algorithm used to manage the health of populations', doi: 10.1126/science.aax2342;

19 'Home | Eticas', (2023), Available at: https://eticas.ai/ (Accessed: 17 May 2024);

20 Jenny Brennan and Alexandru Circiumaru, 'Getting under the hood of big tech', (Ada Lovelace Institute, 15 March 2022), Available at: https://www.adalovelaceinstitute.org/blog/getting-under-the-hood-of-big-tech/ (Accessed: 21 May 2024);

New York City Local Law 144 was one of the first laws to implement an independent algorithm auditing regime

As a practice within the technology sector, algorithm auditing remains uncommon and is used in an ad-hoc manner for some kinds of AI systems.

> The technology sector lacks formal standards of practice, certification schemes or professionalised bodies for how to conduct algorithm audits.

Despite this fact, national governments are increasingly interested in incorporating algorithm auditing requirements for AI systems to serve a variety of goals. These include inspecting AI systems for algorithmic bias against certain demographic groups or compliance with particular laws and regulations.[21] [22] Some regulators in the UK, such as the Competition and Markets Authority or the Information Commissioner's Office, already have the powers to conduct algorithm audits.[23]

However, more recent proposals call for independent auditing regimes of AI systems by a third party. In the USA, several state-level bills have proposed third-party auditing of algorithmic employment decision-making tools (AEDTs).[24] In the EU, the Digital Services Act now requires operators of 'Very Large Online Platforms' (VLOPs) to undergo independent auditing of their content moderation practices. And in the UK, the Government has highlighted that independent bias and compliance audits could comprise part of a wider AI assurance toolkit that may feature in a future AI regulation bill.[25]

21 Rekenkamer, (2022), An Audit of 9 Algorithms used by the Dutch Government - Report - Netherlands Court of Audit, Algemene Rekenkamer, Available at: https://english.rekenkamer.nl/publications/reports/2022/05/18/an-audit-of-9-algorithms-used-by-the-dutch-government (Accessed: 18 January 2024);

22 Mökander, (2023), 'Auditing of AI: Legal, Ethical and Technical Approaches', doi: 10.1007/s44206-023-00074-y;

23 Auditing algorithms: the existing landscape, role of regulators and future outlook, (no date), Available at: https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook (Accessed: 27 September 2023);

24 A look at proposed US state private sector AI legislation, (no date), Available at: https://iapp.org/news/a/a-look-at-proposed-u-s-state-private-sector-ai-legislation (Accessed: 17 May 2024);

25 A pro-innovation approach to AI regulation: government response, (no date), Available at: https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response (Accessed: 17 May 2024);

With many governments and companies looking at how to implement algorithm audits, and to better understand how to implement these regimes effectively, we wanted to explore an early attempt to create an independent algorithm auditing regime. In this project, we sought to study one of the first laws to implement an independent algorithm auditing regime: New York City Local Law 144 (LL 144).

> Taking effect in July 2023, LL 144 requires employers using automated employment decision-making tools (AEDTs) for employment actions, like automatically sifting candidates, to conduct and publish an independent bias audit of their AEDT.

LL 144 is the first attempt to require independent bias testing for commercial machine learning software. This case study provides rich contextual detail around how auditing works in practice, from which we've generated recommendations to inform future algorithm auditing regimes.

This discussion paper is intended as a complement to a research paper co-authored with our research collaborators from Data & Society, which provides more detail on the research background and findings. Read 'Auditing Work: Exploring the New York City algorithmic bias audit regime'.[26]

There is also a companion study by many of the same authors,[27] focused on collecting the published audit reports.

26 Groves L and others, 'Auditing Work: Exploring the New York City Algorithmic Bias Audit Regime' (arXiv, 12 February 2024) http://arxiv.org/abs/2402.08101

27 Wright L and others, 'Null Compliance: NYC Local Law 144 and the Challenges of Algorithm Accountability' https://osf.io/upfdk/ accessed 23 May 2024

How to read this paper

If you're a **researcher** interested in the implementation or evaluation of AI accountability methods, you might focus on the background and context of the emergence of the law, and the 'evidence' sections of the recommendations.

If you're a **policymaker**, you should read the executive summary for an overview of the research project and of our findings, and focus on the 'proposal' sections of our recommendations for solutions for adopting an effective and meaningful auditing regime.

Glossary of key terms

- **A selection rate** is the frequency at which members of a group are chosen to move forward in a hiring/promotion process or rejected/screened out; 'selection' does not refer only to the final hiring decision, but also all decisions before that.

- The law asks for **impact ratio** measures to be presented in the audit report. Impact ratios are a method for measuring discriminatory outcomes as the relative selection rate between demographic groups. The numerator is the selection rate of the least selected/protected group, and the denominator is the selection rate of the most-selected group (or full population). An impact ratio of 1.0 means a perfectly equal selection rate between groups; an impact ratio of less than one indicates a discriminatory outcome against the less-selected group. The lower the fraction, the more discriminatory the outcome is.

- **Disparate impact (also referred to as 'adverse impact')** has emerged from the USA as a convention referring to (e.g. employment) decisions that have an unacceptable exclusionary or discriminatory effect on certain groups (classified by ethnicity, gender, age, etc.) The presence of disparate impact is measured by the outcomes of a decision process, and does not take into account any intent to discriminate.

- **The four-fifths rule** is a conventional way to determine whether there is disparate impact in a selection process. The rule states that the selection ratio of a minority group should be at least four-fifths (80%) of the selection ratio of the majority group. Falling short of the four-fifths rule draws additional regulatory scrutiny and therefore is deeply ingrained in US hiring practices.

- **Automated employment decision-making tool, or AEDT**: a system that

  - uses machine learning, statistical modelling, data analytics, or artificial intelligence, AND

  - helps employers and employment agencies make employment decisions, AND

  - *substantially assists* or replaces discretionary decision-making.[28]

In the law, '**substantially assist**' is defined as being the primary/majority reason, or predominant reason among several, for a hiring decision. To the authors' best knowledge, there is no system on the market that fully replaces human decision-making in hiring, which results in the scope of the law being highly dependent on the interpretation of 'substantially assist'.

28 'Automated Employment Decision Tools: Frequently Asked Questions', (no date);

AI systems may be less effective for genders or ethnicities underrepresented in their training data

Background and context

In the following sections, we present an overview of both the context around use of AI in hiring decisions, and an introduction to New York City's Local Law 144.

AI in hiring

AI is increasingly being used by employers in the process of hiring job applicants. Surveys of employers find that many believe algorithmic systems can streamline processes like CV sifting and recruitment, particularly in scenarios where large numbers of candidates may apply for a role.[29] [30]

> AI systems are used in different phases of the hiring process, from sourcing and screening to the final selection of a candidate.

For example, hiring platforms that post job listings might use algorithms to scan for potential candidates, while the employer posting the job might use AI to review CVs to filter applicants who meet the required criteria or analyse candidates' video interviews for personality traits to determine culture fit.

Some hiring systems use more sophisticated machine learning technologies, such as natural language processing (NLP) or computer vision, which could be used in a personality assessment test[31] or in an interview. These kinds of candidate tests claim to assess a candidate's

29 Pew Research Center, (2023), AI in Hiring and Evaluation of Workers: What People Think, Pew Research Center, Available at: https://www.pewresearch.org/internet/2023/04/20/ai-in-hiring-and-evaluating-workers-what-americans-think/;

30 Workable's AI in Hiring Survey, (no date), Available at: https://get.workable.com/ai-in-hiring-survey (Accessed: 17 May 2024);

31 How Can Hiring Managers Use the OCEAN Personality Test in Recruitment? | Turing, (2022), Available at: https://www.turing.com/blog/ocean-personality-test-for-hiring-interviews/ (Accessed: 17 May 2024);

'willingness to learn' or general aptitude, or even attributes like 'agreeableness'.[32]

Proponents of these tools suggest they can streamline some of the more resource-intensive processes in hiring, freeing up employer time.[33] However, these tools can also exacerbate discrimination toward minoritised groups with protected characteristics.[34] For example, these technologies may be less effective for genders or ethnicities that are underrepresented in the data used to train the system.

Local Law 144

Local Law 144 (LL 144) is the first legally mandated algorithmic bias auditing regime. Under this law, employers and employment agencies using AEDTs must be subject to an independent bias audit. The results of these bias audits must be publicly listed, along with a notice to jobseekers that AEDTs are being used in the hiring process, for all job roles based in New York City.

These obligations amount to mandated transparency requirements. The law only requires companies to make the results of the audit transparent but does not prescribe any particular action after the audit has been completed and posted; it does not prescribe the removal or correction of a biased model.

LL 144 was first introduced in July 2020, as a result of civil society efforts advocating for regulation of algorithmic systems,[35] and passed in late 2021. (See the Appendix for full details of the law drafting process, including revisions and public comment rounds). The City Commission tasked the Department of Consumer and Worker Protection (DCWP) with drafting the specific rules that implement the law.

As a result of the public commenting periods and lobbying efforts on behalf of major employers and major AEDT tool vendors, the text of the proposed rules changed substantially during public comment and revision periods.

32 Demopoulos, (2024), 'The job applicants shut out by AI: "The interviewer sounded like Siri"', Available at: https://www.theguardian.com/technology/2024/mar/06/ai-interviews-job-applications (Accessed: 17 May 2024);

33 Naik, (no date), Council Post: How Artificial Intelligence Benefits Recruiting, Available at: https://www.forbes.com/sites/forbesbusinesscouncil/2023/06/01/how-artificial-intelligence-benefits-recruiting/ (Accessed: 17 May 2024);

34 Sloane, (2021), 'The Algorithmic Auditing Trap', Available at: https://onezero.medium.com/the-algorithmic-auditing-trap-9a6f2d4d461d (Accessed: 29 June 2023);

35 Cahn, (2021), 'New York City's Surveillance Battle Offers National Lessons', Available at: https://www.wired.com/story/opinion-new-york-citys-surveillance-battle-offers-national-lessons/ (Accessed: 18 January 2024);

We highlight some of the key changes below:

- **Changes to key definitions:** by the time of implementation, the law had widened the definition of an AEDT to include systems that only 'substantially assist or replace' hiring decisions, *limiting the number of systems in scope of the law.*

- **Changes to scope of the law:** at the outset, vendors and developers of AEDT would be subject to bias audit: this later changed to employers and employment agencies – 'the end users' – *changing the accountability relationships.*

One interviewee we spoke to, at a company with knowledge of the trajectory of the lawmaking, said:

> 'Because of the, frankly, quite heavy lobby from the larger employment companies, the scope has significantly narrowed from what we believe the original intent of the law was.' – Interviewee offering a 'pre-audit' service

Our research and methodology

This project spanned from June 2023 to January 2024 as a partnership with Data & Society to explore what lessons could be learned from the NYC algorithmic bias audit law for other governments implementing similar schemes.

We conducted 17 interviews with 15 practitioners and experts offering audits in the regime or with knowledge of the law drafting process to explore the following research questions:

- **RQ1:** What are the practical components of a bias audit in this context?

- **RQ2:** What are the components, relationships and incentives that make for an effective bias auditing regime?

- **RQ3:** What are the experiences of auditors, and how can we use those experiences to inform wider policy and practice around other algorithm auditing regimes?

Our findings reveal some modest successes under this regime, including the creation and use of a standardised test for bias auditing, and reports that companies subject to this law adopted wider responsible AI and ethical data practices that they might not have otherwise adopted.

> Overall, LL 144 failed to meet its objectives of curbing unjust hiring practices and reducing the use of discriminatory AEDTs.

One interviewee shared their view about how LL 144 could be iterated on and developed further in future lawmaking:

'I don't think [LL 144] should be admonished for not being perfect. I call this the first pancake: without the first pancake, none of the other pancakes would be better. And the first one is always pretty awful in the pan, no matter what you do.' – Interviewee offering a 'pre-audit' service

Policymakers will need to consider how audits are designed, operationalised and evaluated

Recommendations

The six main recommendations we present in this chapter are drawn from the research and interview findings of this project, and are primarily aimed at policymakers interested in developing and mandating audit regimes for AI.

Each recommendation is set out according to the following structure:

- **The challenge:** defines and explains the challenge needing to be addressed.
- **The evidence:** provides an overview of the research evidence that contributed to our understanding of the challenge, where we draw from our interviews and an analysis of the law.
- **The proposal:** puts forward the policy solution(s) to the challenges identified.

An effective auditing regime will require a wide focus, with policymakers needing to consider elements beyond just how the audit is designed, to examine how it is operationalised and evaluated. Accordingly, our recommendations offer policy proposals across:

- **methodological components of audit and auditing practice** (e.g. what metrics should be adopted, what standards of practice created?)
- **the design and operation of the regime** (e.g. how should outcomes be measured?)
- **compliance with, and enforcement of, the regime** (e.g. how can audits result in meaningful enforcement action?).

Recommendation 1: Auditing laws must establish clear definitions that clearly capture the full range of AI systems in scope

The challenge: an inadequately defined scope can make an audit regime less effective

Policymakers seeking to establish an algorithm auditing regime must address several preliminary questions. These include defining what AI systems should be audited, how that audit should be conducted (including the components of the audit) and by whom. There are different motivations behind why policymakers might want to use or mandate an algorithm audit, from evaluation or scrutiny of AI-derived decisions or outputs[36] to assessing organisational compliance with AI policy.[37]

It is also important to establish what the intended target of the audit is (what are you checking for?) and the success criteria (what is your goal, and how will you know when you've met it?). Forthcoming research from the Ada Lovelace Institute on evaluations of advanced AI systems[38] finds that defining and answering these questions is incredibly difficult, especially when considering the range of risks that AI systems might raise.

The evidence

Our findings show that LL 144 lacks clarity on which AI systems were in scope. Several key definitions of LL 144 were changed over the lawmaking process due to industry lobbying and made more open to interpretation by companies who are subject to this law. According to our interviews, this created loopholes that significantly limit the number of both systems – and therefore employers – in scope. The law says only AEDTs that 'substantially assist or replace' human decision-making are in scope:

36 Raji et al., (2022), 'Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance', arXiv, Available at: http://arxiv.org/abs/2206.04737 (Accessed: 4 August 2022);

37 A Guide to ICO Audit Artificial Intelligence (AI) Audits', (no date);

38 'Evaluation of Foundation Models' https://www.adalovelaceinstitute.org/project/evaluation-foundation-models/ accessed 29 May 2024

> 'The 'substantially assisting' [per the definition of an AEDT] definition 'creates huge loopholes. Because you can say, hey, there's always a human that ultimately clicks something' – Interviewee speaking on background to the law'

'[on the 'substantially assisting' definition] And we kind of saw clients and prospective clients, say "well, that [doesn't apply to] me". But from our point of view, it is still important to do the audit. Because even if the system is using early phases, and it's not the most important decision, it can still have these downstream effects.' – Interviewee from an audit company

The proposal: Involve a wide selection of actors to determine which technologies are in scope

Lawmakers designing auditing laws, metrics and standards must adopt a participatory process that involves perspectives of people affected by AI systems, as well as those of civil society groups.[39] This will help ensure that the appropriate range of AI systems are captured in scope of auditing laws. For example, for auditing regimes around AEDTs, lawmakers should consult closely with union representatives, workers and advocacy groups, and perhaps even include these representatives in a body establishing standards of practice and which AI systems should be in scope.

The importance of a collaborative standards-setting exercise with input from a broad range of actors cannot be overstated. This is critical to help collectively define the kinds of risks that AI will raise, like 'systemic risk,' and develop metrics for these tests.[40]

Policymakers should also be prepared to take an iterative approach to metrics and standards development, which may need adapting in light of new empirical evidence on the effectiveness of auditing in practice.

39 Meaningful public participation and AI, (no date), Available at: https://www.adalovelaceinstitute.org/blog/meaningful-public-participation-and-ai/ (Accessed: 17 May 2024);

40 Ada Lovelace Institute, (2023), Inclusive AI governance: Civil society participation in standards development, Available at: https://www.adalovelaceinstitute.org/report/inclusive-ai-governance/;

Recommendation 2: Auditing laws must establish clear standards of practice on the role and responsibilities of auditors

The challenge: a lack of clarity on roles and practices creates conflicts of interests

Unlike industries such as financial services, there is currently no formalised industry of independent algorithm auditors. Several bodies and organisations have designed audit procedures and standards, and there is an increasing market of companies who offer 'audit as a service' in addition to their main business offering. This lack of clarity raises serious challenges for determining whether an audit is truly 'independent'.

The evidence

LL 144 ran into challenges on **who should conduct an audit and who is audited**. The law applies to employers using AEDTs, but not to vendors who build these tools and sell them. It defines independent auditors as third-party experts who have no financial stake in the success of the product or the financial outcome of the employer. However, our interviews reveal a variety of different organisations (including data analytics companies, law firms and AI governance startups) who interpreted this criteria and their roles in different ways. These companies offered four different kinds of audit services:

- **Companies offering a 'pre-audit' service to help auditees get 'audit ready'.**
- **Companies conducting the audit and writing the audit report for a client.**
- **Companies offering additional guidance and mitigation strategies to auditees.**
- **Companies offering a service to certify that an audit has been conducted in an appropriate manner.**

A 'pre-audit' service, according to our interviews, might include helping employers streamline their data collection and provenance processes, or provide information data governance tools. Auditors offering this service that we spoke to did not go on to conduct the audit:

> 'I think it's important that we are not auditors, because auditors by our interpretation of that definition, are an independent authority [. . .] nothing that an auditor does should be taken as advice. We see them as an assessor that can validate, because that's the primary function of an auditor.'
> – Interviewee offering a 'pre-audit' service

The majority of our interviewees were with companies offering the bias audit as directed under LL 144, and preparing the audit report for the client (See the 'Appendix' for an example final audit report). Some companies also offered additional guidance and mitigation strategies, often as part of a wider service delivery (for example, responsible AI consulting).

Finally, several companies offered a service to 'certify' LL 144 audits to assess auditing conduct or the results. Many companies offering this service felt the need to take on this role due to LL 144 not requiring action from the results of the bias audit.[41]

The diversity of auditors and roles identified in this research signals promising potential for the emerging 'audit as a service' market, but without standards or accreditation, there is a risk that consistency and rigour of algorithm audit services will vary hugely.

The proposal: develop standardised audit practices and auditor oversight bodies

Independent algorithm auditing regimes require auditors to follow clearly defined practices that represent a high degree of rigour and credibility.[42] These regimes involve the delegation of authority to an independent body to conduct risk assessments. It is crucial that policymakers create the necessary standards of practice and oversight regimes to ensure algorithm auditors follow approved practices.[43] An oversight body must have the

41 'Automated Employment Decision Tools: Frequently Asked Questions', (no date);

42 AI Accountability Policy Report | National Telecommunications and Information Administration, (no date), Available at: https://www.ntia.gov/issues/artificial-intelligence/ai-accountability-policy-report (Accessed: 17 May 2024);

43 Birhane et al., (2024), 'AI auditing: The Broken Bus on the Road to AI Accountability', arXiv, Available at: http://arxiv.org/abs/2401.14462 (Accessed: 31 January 2024);

powers, skills and resources to certify and check the work of independent auditing agencies and update standards of practice when necessary.[44]

Standards should also acknowledge the different roles within an auditing regime – such as organisations that help get companies audit-ready and organisations that certify if an audit has taken place. These roles should be kept separate, with clear standards drawn for each of these roles.

Sectors that frequently make use of third-party auditing, such as the financial sector, place strict guardrails around the degree of closeness an auditor can have with an auditee.[45] This kind of relationship dynamic should also apply to algorithm auditing to avoid companies 'marking their own homework'. For example, the financial services approach of accrediting third-party auditors could be a replicated, with certified algorithm auditors appearing on a public register (similar to the Institute of Chartered Accountants in England and Wales register).[46]

To start this process, policymakers could seek to support an emerging audit ecosystem by convening and disseminating best practices as they exist now. For example, the UK Responsible Technology Adoption Unit has already published case studies of AI assurance techniques in use across various sectors,[47] outlining approaches taken and the benefits of using the technique for the organisation. The UK Government could consider expanding these case studies to spotlight companies offering algorithm audit and related services to outline the functionality and importance of the different auditor roles. Any independent oversight body or standards of practice must be created in close partnership with civil society organisations and affected communities to ensure fair representation of their interests.

44 Raji et al., (2022), 'Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance', arXiv, Available at: http://arxiv.org/abs/2206.04737 (Accessed: 8 August 2022);

45 Raji et al., (2022), 'Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance', arXiv, Available at: http://arxiv.org/abs/2206.04737 (Accessed: 8 August 2022);

46 Audit registers, (no date), Available at: https://www.icaew.com/library/subject-gateways/auditing/audit-registers (Accessed: 17 May 2024);

47 Portfolio of AI assurance techniques - GOV.UK, (no date), Available at: https://www.gov.uk/guidance/portfolio-of-ai-assurance-techniques (Accessed: 17 May 2024);

Recommendation 3: Auditing laws must enable smooth data collection for auditors

The challenge: auditors struggle to gain access to the right levels of data from companies they are auditing

Auditing AI systems is not so much a technical challenge as a relational one. A primary job of an auditor is to access the relevant data and conduct a test. While bias audits under LL 144 were technically simple to conduct, auditors routinely struggled to obtain the data needed to conduct the audit. Some of these challenges were cultural, with companies refusing to acknowledge they may be using biased tools. Other challenges related to companies failing to track the data necessary to conduct an audit.

To conduct the tests for an audit, auditors need access to sufficient amounts of the relevant data. There is existing evidence of companies and platforms withdrawing the APIs (application programming interfaces) that enable third-party auditor access.

The evidence

Auditors often needed to mediate relationships with both the employer (audited organisation) and the AEDT developer or vendor, as the law permits use of hiring-rate data owned by the vendor if the employer did not have the optimal amount of data stored themselves. One auditor shared their experience:

‘Our only adversarial or third party was the vendor. So, we had to kind of take a very hard stance, getting our position and be kind of aggressive with the vendor to get us the data that we wanted so that we could produce that audit the way it’s meant to be’ – Interviewee previously employed by an audit company

In our interviews, many auditors described difficulty in obtaining demographic data, finding that many employers do not hold information such as ethnicity and gender of applicants to their job postings. In absence of sufficient demographic data, some auditors reported making inferences about gender and ethnicity from the candidate’s name, a problematic practice in diverse urban areas such as New York City (as well as being a prohibited practice under LL 144).

'The company we were doing [the audit] for didn't have [demographic data]. And they just said, well, the [gender split] is about 50/50 in the US. And so, they just assumed 50/50 across the board for all their selection rates' – Interviewee from an audit company

Additionally, auditors reported surfacing thorny cultural attitudes to data at some audited employers. This included expressions of discomfort at the suggestion they could be making biased decisions:

> '[Some companies] say "we don't collect either on sex and ethnicity, and therefore we can't be biased", and it's mistaken, but it can be very powerful' – Interviewee from an audit company

These findings are good examples of complex dynamics of policy solutions like algorithm audits. Policymakers need to be aware of how companies being audited are likely to respond and what their pain points will be, and design laws that enable companies to have a clear obligation to provide data to an independent auditor and comply.

The proposal: auditing laws must mandate the appropriate levels of data access for an auditor

Policymakers need to design laws that require companies to provide data to an independent auditor. Auditing laws need clear language around permitted data collection methods and tools to facilitate auditor work. An effective audit regime also requires robust underlying data governance processes to accompany the outlined methods for auditing.

Policymakers should consider setting standards for data cleaning to ensure companies are actively managing inconsistencies and duplicates to improve regularity. Additionally, guidance should also be provided for companies that wish to comply with the audit mandate but don't collect demographic data. For example, shared secure data stores hosted by AEDT vendors could supply companies with sufficient data, with mandates for creating data access.

Recommendation 4: Auditing laws must establish meaningful metrics that accurately capture a risk

The challenge: it can be difficult to identify a metric to audit for that accurately reflects a risk

For policymakers seeking to create algorithm auditing regimes, one major challenge will be to identify a clear test and metric that accurately captures the level of acceptable risk for an AI system.

In LL 144, the prescribed method of bias auditing is an assessment of demographic parity (fairness) across gender and race/ethnicity, as well as the intersection between these categories, to produce three impact ratios (see definitions above). These ratios act as a proxy for determining if a system is unlawfully biased. Auditors are not required to test systems for other forms of bias, including against ability and age.

The evidence

LL 144 won praise from some of our interviewees for a straightforward audit procedure that was thought to help effectively convey the parameters of the assessment to auditors. The law does not prescribe its own metric for success for auditing, but many auditors followed the 'four-fifths' convention (see the Glossary) a widely known standard set by the US's Equal Employment Opportunity Commission (EEOC).

> 'I would say that it has been helpful as something that's inspired debate about metrics, because it's one of the few policies that actually defined a metric, you know, the disparate impact analysis.' – Interviewee offering a 'pre-audit' service

The proposal: develop region-specific and risk-specific audit metrics, and use other methods where a metric will not accurately capture a risk

The bias auditing regime under New York City LL 144 is regionally specific and the metrics are tied to US employment and anti-discrimination law. UK and EU policymakers will need to generate jurisdiction-specific metrics for algorithmic bias.

In the UK, the Information Commissioner's Office has created guidance for algorithm auditing.[48] However, this guidance needs to be extended to include metrics for tests for bias auditing. This could include the Alan Turing Institute's categories for individual and group fairness metrics.[49] It must also account for intersectional forms of bias where there may be multiple kinds of protected characteristics occurring at once (that is, going beyond just race and ethnicity categories required in LL 144). This might comprise extending the Public Sector Equality Duty – which public authorities must comply with – to private companies.

Policymakers will need to create similar metrics, standards and practices for auditing for other kinds of algorithmic risks like illegal content, the prevalence of false information and environmental impacts. In the context of auditing algorithms used on social media platforms, the European Union's Digital Services Act[50] sets out categories of systemic risk of AI systems, such as the dissemination of illegal content, but researchers have noted that, in this context, it is unclear when a risk becomes 'systemic'.

It is crucial to acknowledge that some risks cannot be quantified as a clear metric. This will require policymakers to establish other mechanisms, such as citizen review boards, to develop agile thresholds to determine appropriate AI uses in certain contexts. It will also require policymakers to consider what burden of proof must be met for some AI systems to be allowed market entry.

In some high-risk contexts, AI systems that cannot be demonstrably evaluated for certain risks may need to be prohibited. There is precedence for this approach in finance, where the US Consumer Financial Protection Bureau has prohibited the use of 'black box' AI systems used to determine loan application decisions because they cannot be easily audited for their discriminatory impacts.[51]

48 'Guidance on the AI auditing framework Draft guidance for consultation', (no date);

49 AI Ethics and Governance in Practice: AI Fairness in Practice, (no date), Available at: https://www.turing.ac.uk/news/publications/ai-ethics-and-governance-practice-ai-fairness-practice (Accessed: 17 May 2024);

50 European Parliament and Council of the European Union, (2022), 'Digital Services Act', Official Journal of the European Union;

51 CFPB Acts to Protect the Public from Black-Box Credit Models Using Complex Algorithms, (2022), Available at: https://www.consumerfinance.gov/about-us/newsroom/cfpb-acts-to-protect-the-public-from-black-box-credit-models-using-complex-algorithms/ (Accessed: 17 May 2024);

Ultimately, auditing under LL 144 does not create accountability between candidates and employers

Recommendation 5: Audits should follow a theory of change that results in meaningful outcomes

The challenge: what is an audit set up to achieve?

As an accountability mechanism, the goal of algorithm audits should always be to ensure the developers of AI systems are held accountable when their systems fail or break, to prevent or remove failed products from the market, and to enable modes of redress when harm has occurred.

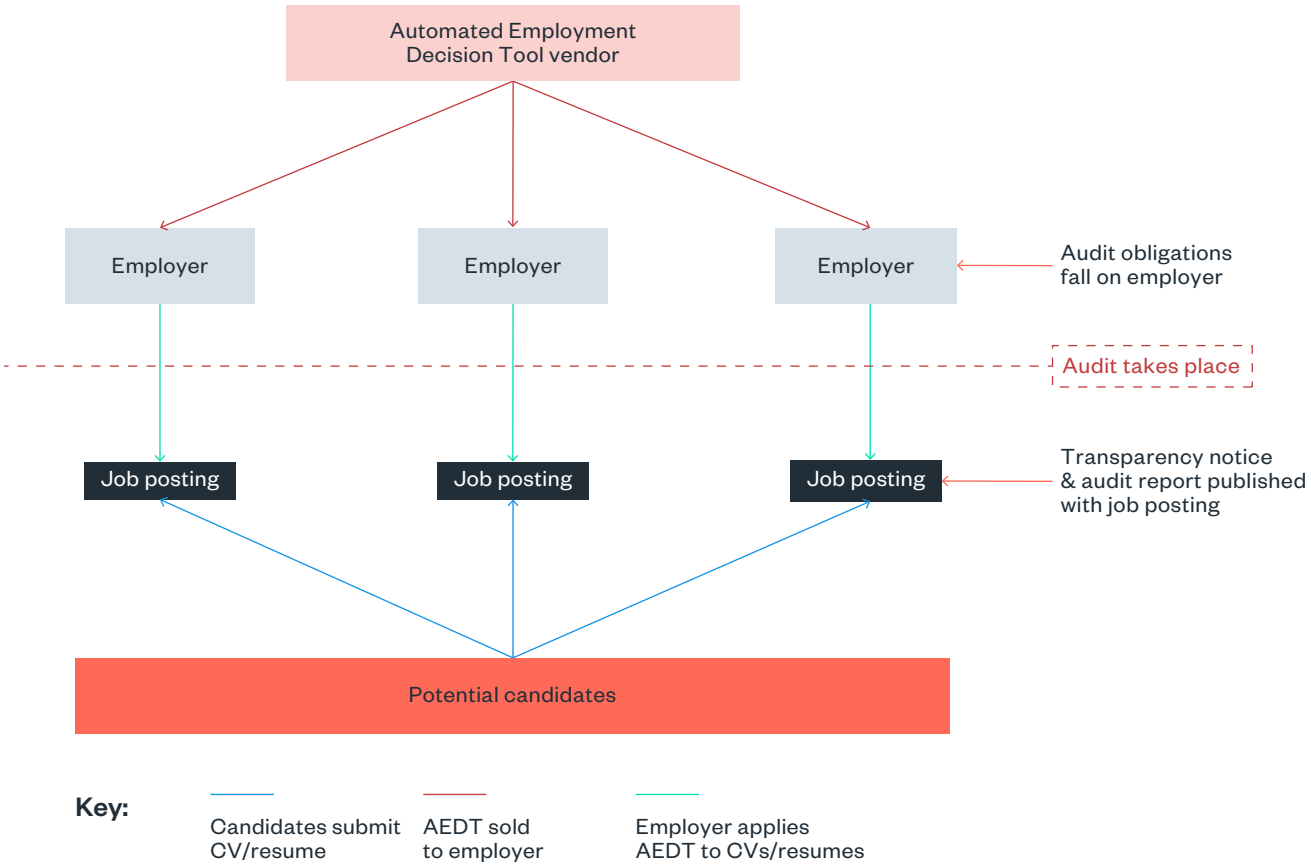
The evidence

In its regulatory design, LL 144 failed to create adequate accountability dynamics. It adopts a transparency-driven theory of change, theorising that the publication of completed audits would provide New York City candidates some choice about whether to be subject to an AI system in the hiring process. The law does not require companies to stop using an AI system that is demonstrated to display bias, nor does the law provide a clear legal remedy for individuals who experience algorithmic bias from these tools.

The intent behind the implementation of LL 144, according to the law's sponsors, was to 'curb unjust practices in hiring'.[52] Due to lobbying efforts, the target of the law changed from placing obligations on AEDT vendors (developers), to the employers (the 'end users'). This has implications for accountability, as it means AEDT vendors are not prevented from selling potentially biased or unsafe products under this law.

52 Ivanova, (2020), New York City wants to restrict artificial intelligence in hiring - CBS News, Available at: https://www.cbsnews.com/news/new-york-city-artificial-intelligence-hiring-restriction/ (Accessed: 17 May 2024);

Figure 1: New York City Local Law 144 obligations across the supply chain



Elements of the law (the requirement to publish the completed audit report and give notice to candidates about AEDT usage) are designed to create transparency for candidates. This then enables candidates to choose for their application to be reviewed by a human and not be subject to an AEDT. However, candidates may struggle to parse the meaning or significance of an audit (displayed in a statistical table – see the Appendix) and may experience concern that forgoing an AEDT may disadvantage them with a potential employer.

Auditors also noted that because LL 144 only requires auditors to assess for bias across two protected categories – race and gender (and the intersection between), and excludes categories such as disability and age – many potential candidates would not be protected under the law. This evidence highlights the challenge around setting parameters of an assessment, which may result in exclusionary outcomes for some groups at the expense of others.

Ultimately, auditing under LL 144 does not create accountability between candidates and employers. If a candidate wished to challenge whether they were unfairly discriminated against using an AEDT, they would need to pursue plaintiff litigation, where they could use the audit as evidence. But this is a higher bar and creates a significant onus for action on job candidates.

Our interviews also revealed that auditors themselves do not take on the role of ensuring job candidates are protected from risk. Instead, auditors are focused on the task of conducting or checking the audit for their client – either an employer who has procured an AEDT, or a vendor building a tool.

'The needs of the [candidate] are so different from the needs of an enterprise [. . .] and although we care about the [candidate], I don't think that it is incumbent upon [us] to be able to satisfy the needs of that [candidate] at the same time as satisfying the needs of the enterprise.' – Interviewee from a company offering a 'pre-audit' service

The proposal: Audits should be made transparent, publicly accessible and legible for lay audiences via a transparency register

As the failures of LL 144 show, it is insufficient to solely rely on transparency to create accountability. However, to enable the beneficiaries of an audit regime to scrutinise, evaluate and contest its outcomes, there is a need for policymakers to establish clear audit reporting that makes the audit results clear for non-technical audiences.

This could include a summary of results in simple language and an outline of potential next steps, as well as named contact details on audit reports (following, for example, the UK Government's 'two tier' transparency approach adopted in the UK algorithmic transparency standard for public-sector AI tools).[53] Central governments should also establish and host an audit repository or AI risk register, like those used in post-market monitoring of medical devices.[54]

53 Algorithmic Transparency Recording Standard - Guidance for Public Sector Bodies, (no date), Available at: https://www.gov.uk/government/publications/guidance-for-organisations-using-the-algorithmic-transparency-recording-standard/algorithmic-transparency-recording-standard-guidance-for-public-sector-bodies (Accessed: 17 May 2024);

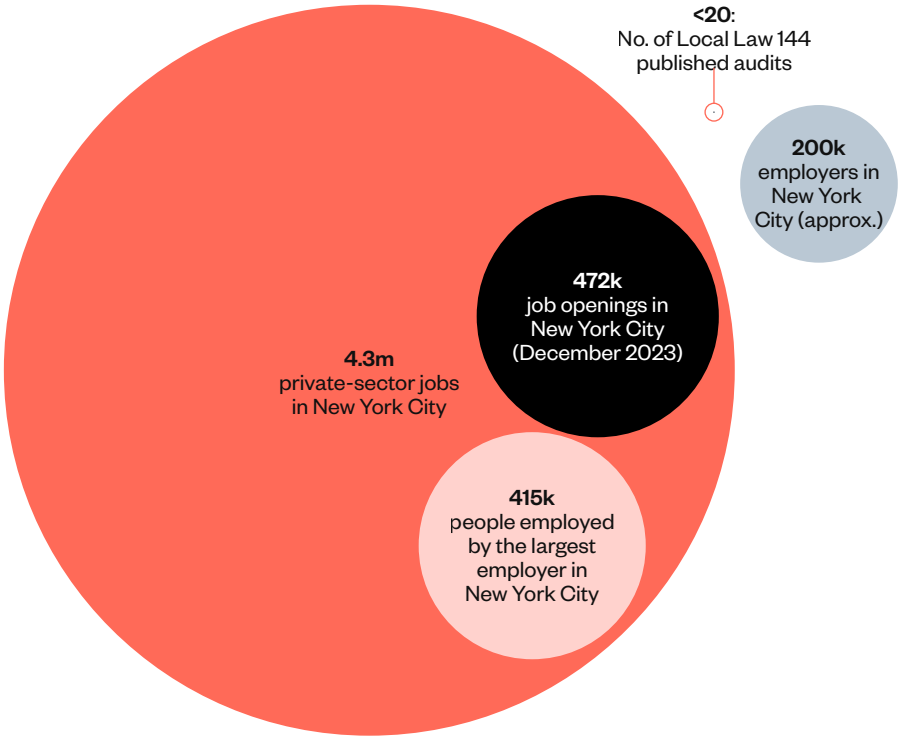
54 Guidance for post-market surveillance and market surveillance of medical devices, including in vitro diagnostics, (no date), Available at: https://www.who.int/publications/i/item/9789240015319 (Accessed: 17 May 2024);

Recommendation 6: Auditing laws need mechanisms to monitor and enforce against non-compliance

The challenge: creating a regime with adequate penalties for non-compliance

Successful governance regimes require strong enforcement mechanisms to incentivise compliance. LL 144 failed to create a meaningful penalty to incentivise companies to comply and publish its audits. The law provides for fines of between $500 and $1,500 per day for non-compliance, meaning the lack of transparency notice and audit report. As of March 2024, over six months after the law was enacted, there are only 20 published audits available online from the over 200,000 employers in New York City (some originally published audits have disappeared).[55] The very low level of compliance demonstrates that the penalties were not substantial enough to instigate action.

Figure 2: Coverage of LL 144 audits in the NYC job market[56]



55 OSF | Null Compliance: NYC Local Law 144 and the Challenges of Algorithm Accountability, (no date), Available at: https://osf.io/upfdk/ (Accessed: 24 January 2024);

56 'Labor Statistics for the New York City Region' (Department of Labor) https://dol.ny.gov/labor-statistics-new-york-city-region; 'Employment Data | New York City by the Numbers' (26 October 2021) https://ibo.nyc.ny.us/cgi-park2/category/employment/; 'New York Job Openings and Labor Turnover — December 2023 : Northeast Information Office : U.S. Bureau of Labor Statistics' (Bureau of Labor Statistics) https://www.bls.gov/regions/northeast/news-release/2024/jobopeningslaborturnover_newyork_20240221.htm; 'These Are The 100 Largest Companies In New York' (Zippia, 19 March 2024) https://www.zippia.com/advice/largest-companies-in-new-york/; Wright L and others, 'Null Compliance: NYC Local Law 144 and the Challenges of Algorithm Accountability' https://osf.io/upfdk/ accessed 23 May 2024

The evidence

Under LL 144, the decision about whether to carry out an audit and publish the report rests with the employer using the AEDT. For example, a potential scenario where an audit revealed disparate impact would potentially have wider ramifications in terms of breaching federal-level employment discrimination code as set out by the USA's Equal Employment Opportunity Commission (EEOC):

> 'Everybody is scared about putting anything out that's below point eight [0.8 or 'four-fifths', referring to the 'four-fifths' convention or metric for determining bias in US employment law – see Glossary above]. It's basically grounds for a lawsuit because now you know that they are not adhering to this guideline' – Interviewee with experience at an AEDT vendor

Some auditors said that some clients preferred to wait and see whether the penalties for instances of non-compliance would be enforced by the regulator, the New York City Department of Consumer and Worker Protection (DCWP). A complementary research study led by Data & Society and Cornell University Citizen and Technology Lab conducted a mapping of the publicly available audit reports.[57] As their findings show, the lack of audit or notice on an employer's website does not necessarily indicate non-compliance, as an employer may not be using an AEDT at all, or may have determined that their tool is out of scope. It also does not indicate that an audit has not taken place: our interviews revealed instances of auditors preparing audits for clients, who chose not to publish the reports. The researchers called this phenomenon 'null compliance'.

Additionally, some auditors expressed a desire to help both clients and vendors update processes that might improve sub-par audit results – but beyond the requirement for completing the audit again, biannually, there is no feature of the law that helps facilitate this.

57 OSF | Null Compliance: NYC Local Law 144 and the Challenges of Algorithm Accountability, (no date), Available at: https://osf.io/upfdk/ (Accessed: 24 January 2024);

Our interviews also revealed scepticism among some auditors about the capacity of the regulator to properly enforce the law, with implications for compliance:

'You know, we talk about GDPR, and enforcement of GDPR being spotty in and of itself. And that's a very far-reaching regulation that's had global implications. And if, if the EU is struggling with compliance, and I'm sure the DCWP is also struggling with compliance.' – Interviewee at a 'pre-audit' company

'Enforcement is not fantastic, I would say. And again, I think that's partly because of the administration. They're trying to be business friendly – they are not enforcing it as they should.' – Interviewee with experience at an AEDT vendor

The proposal: Audits should lead to meaningful and serious enforcement action

Policymakers creating algorithm auditing regimes must ensure these regimes lead to products being denied entry to the market, or being removed from the market if they fail to comply with local laws. This could be done by using audits as a pre-market entry requirement that specifies that a developer's AEDT or system could not be sold unless it met the specified metric or test.

Regular audits following market entry would ensure biased systems could be removed from the market. Policymakers must impose serious sanctions to incentivise companies to comply with auditing regimes. Regulators will require inspection and enforcement powers to mandate these tests and to fine bad actors who continue to operate faulty or illegal systems. This includes ensuring regulators are sufficiently resourced to hire staff to conduct enforcement operations.

One model that could be adopted is model used by the US Food and Drug Administration (FDA), which requires medical device and drug manufacturers to use pre-market audits and post-market monitoring to ensures that faulty or unsafe products are removed from the market.[58]

58 Ada Lovelace Institute, Safe before sale, (2023), Available at: https://www.adalovelaceinstitute.org/report/safe-before-sale/ (Accessed: 17 May 2024);

Conclusion and further questions

This project provides important empirical evidence around practical considerations and dynamics that auditors face.

Competent audit regimes are critical to ensuring meaningful and effective governance of AI systems. Audits are not a complete solution for creating safe and ethical systems but should be adopted as part of a broader toolkit of practices that create accountability between developers of AI systems and the impacted people.

While this research concludes that the bias auditing regime in LL 144 was not effective, our interviews reveal that there are promising avenues for further research and policy development around audit regimes.

The diversity of different roles, functions and expertise that auditing has attracted in the LL 144 points to the potential for a flourishing algorithm auditing market, and there are opportunities for policymakers to help grow and shape this market. However, our findings also lay bare some of the risks around an ineffectual auditing regime: most notably, around failing to create accountability and failing to deliver positive outcomes for people and society.

Many of our interviewees noted that, as the first mandate of its kind, LL 144 was unlikely to achieve total success. Our research shines a light on some of the key points of tension in this law. This not only generates context-specific lessons for New York City policymakers, but also signals areas of focus that could be applied to auditing regimes more widely.

Further questions

We highlight several remaining questions and considerations, which have wide applicability to AI governance. These themes and findings will be explored in forthcoming research by the Ada Lovelace Institute.

- What other infrastructure and policy mechanisms are required for ensuring algorithm audits lead to more accountable/safer outcomes for those impacted by these systems?
- What is needed for audits of AI systems to function effectively? What specific guidance or requirements do auditors need to do their job well?
- How can policymakers and regulators establish their own criteria and metrics for acceptable risks? In the same way the USA has established its four-fifths rule to measure adverse impact, how can the UK establish its own threshold for unacceptable levels of bias?
- What other mechanisms beyond audit and evaluations are needed for a flourishing AI governance ecosystem?

Appendix

Figure 3: History of NYC Local Law 144

This graphic provides a timeline overview of LL 144, from ideation to implementation.

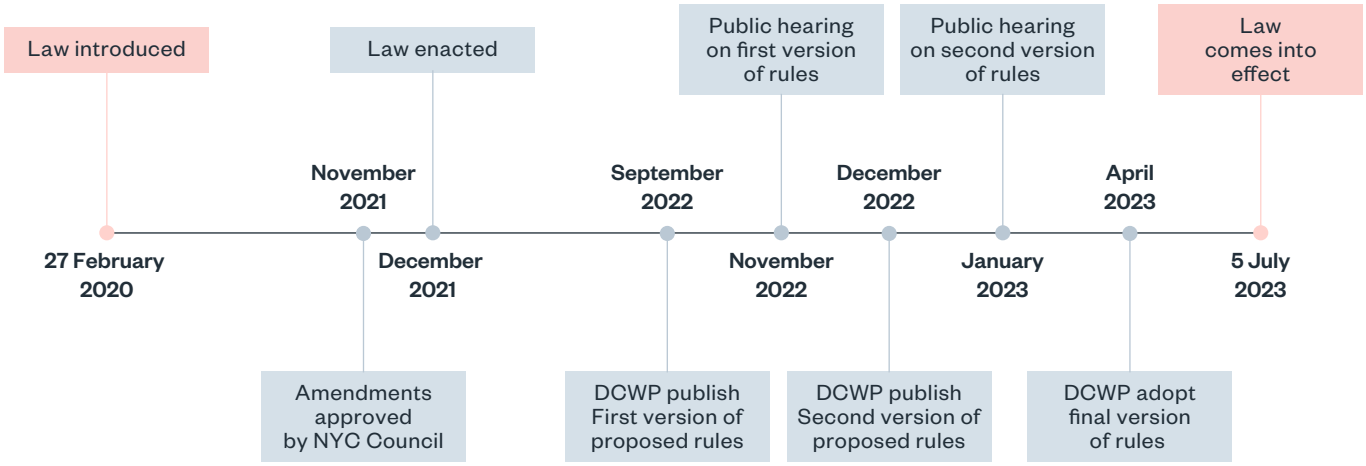


Figure 4: Example audit report, prepared for NBC by ConductorAI[59]

Audit Results

Sex Categories

	# Of Applicants	Scoring Rate	Impact Ratio
Male	2,063,618	47%	0.99
Female	2,007,072	47%	1

Race/Ethnicity Categories

	# Of Applicants	Scoring Rate	Impact Ratio
Hispanic or Latino	470,904	45%	0.89
White	1,457,444	46%	0.9
Black or African American	796,447	48%	0.96
Native Hawaiian or Pacific Islander	14,904	45%	0.9
Asian	941,554	50%	1.00
Native American or Alaska Native	29,612	43%	0.85
Two or More Races	216,700	46%	0.92

Intersectional Categories

	Sex	Race/Ethnicity Categories	#Of Applicants	Scoring Rate	Impact Ratio
Hispanic or Latino	Male		231,334	44%	0.87
Hispanic or Latino	Female		237,315	45%	0.89
Non-Hispanic or Latino	Male	White	725,976	46%	0.9
Non-Hispanic or Latino	Male	Black or African American	351,245	48%	0.94
Non-Hispanic or Latino	Male	Native Hawaiian or Pacific Islander	7,656	45%	0.89
Non-Hispanic or Latino	Male	Asian	547,776	51%	1
Non-Hispanic or Latino	Male	Native American or Alaska Native	17,000	41%	0.82
Non-Hispanic or Latino	Male	Two or More Races	99,271	46%	0.91
Non-Hispanic or Latino	Female	White	723,288	46%	0.91
Non-Hispanic or Latino	Female	Black or African American	442,836	49%	0.97
Non-Hispanic or Latino	Female	Native Hawaiian or Pacific Islander	7,117	46%	0.90
Non-Hispanic or Latino	Female	Asian	391,030	50%	0.99
Non-Hispanic or Latino	Female	Native American or Alaska Native	12,262	44%	0.88
Non-Hispanic or Latino	Female	Two or More Races	113,712	47%	0.93

59 ConductorAI, (no date), Available at: https://www.conductorai.co/nyc-144-audits/smartassistant (Accessed: 17 May 2024);

Acknowledgements

This paper was lead authored by Lara Groves with substantive input from Andrew Strait, Jake Metcalf, Alayna Kennedy and Briana Vecchione.

We would like to thank all our interview participants for their time and expert contributions to this study, both as individuals and representatives of organisations:

- Merve Hickok, Alethicist.org
- Ryan Carrier, ForHumanity
- Yiannis Kanellopoulos, Code4Thought
- Paul White, Resolution Economics
- Evi Fuelle, CredoAI
- Amin Rasekh, CredoAI
- Adrian Byrne, Idiro Analytics
- Ivan Caffrey, Idiro Analytics
- Zachary Goldberg, Trilateral Research
- HolisticAI
- Jey Kumarasamy, Luminos
- Shea Brown, BABL
- Khoa Lam, BABL
- Matthew Boutte, SolasAI
- Trey Causey

And those who preferred not to be named.

About the Ada Lovelace Institute

The Ada Lovelace Institute was established by the Nuffield Foundation in early 2018, in collaboration with the Alan Turing Institute, the Royal Society, the British Academy, the Royal Statistical Society, the Wellcome Trust, Luminate, techUK and the Nuffield Council on Bioethics.

The mission of the Ada Lovelace Institute is to ensure that data and AI work for people and society. We believe that a world where data and AI work for people and society is a world in which the opportunities, benefits and privileges generated by data and AI are justly and equitably distributed and experienced.

We recognise the power asymmetries that exist in ethical and legal debates around the development of data-driven technologies, and will represent people in those conversations. We focus not on the types of technologies we want to build, but on the types of societies we want to build. Through research, policy and practice, we aim to ensure that the transformative power of data and AI is used and harnessed in ways that maximise social wellbeing and put technology at the service of humanity.

We are funded by the Nuffield Foundation, an independent charitable trust with a mission to advance social well-being. The Foundation funds research that informs social policy, primarily in education, welfare and justice. In addition to the Ada Lovelace Institute, the Foundation is also the founder and co-funder of the Nuffield Council on Bioethics and the Nuffield Family Justice Observatory.

Find out more:

Website: Adalovelaceinstitute.org
Twitter: @AdaLovelaceInst
Email: hello@adalovelaceinstitute.org

Preferred citation: Ada Lovelace Institute, *Code & conduct: How to create third-party auditing regimes for AI systems* (2024)
https://www.adalovelaceinstitute.org/report/code-conduct-ai/

ISBN: 978-1-7395236-2-6