Policy Paper

November 2022

AI Audit-Washing and Accountability

Ellen P. Goodman and Julia Tréhu

GMF
Ideas.
Leadership.
Hope.

75 The Marshall Plan 1947–1951
GMF 50

Summary

We are still some distance from a worldwide robot takeover, but artificial intelligence (AI)—the training of computer systems with large data sets to make decisions and solve problems—is revolutionizing the way governments and societies function. AI has enormous potential: accelerating innovation, unlocking new value from data, and increasing productivity by freeing us from mundane tasks. AI can draw new inferences from health data to foster breakthroughs in cancer screening or improve climate modeling and early-warning systems for extreme weather or emergency situations. As we seek solutions to today's vexing problems—climate disruption, social inequality, health crises—AI will be central. Its centrality requires that stakeholders exercise greater governance over AI and hold AI systems accountable for their potential harms, including discriminatory impact, opacity, error, insecurity, privacy violations, and disempowerment.

In this context, calls for audits to assess the impact of algorithmic decision-making systems and expose and mitigate related harms are proliferating, accompanied by the rise of an algorithmic auditing industry and legal codification. These are welcome developments. Audits can provide a flexible co-regulatory solution, allowing necessary innovation in AI while increasing transparency and accountability. AI is a crucial element of the growing tech competition between authoritarian and democratic states—and ensuring that AI is accountable and trusted is a key part of ensuring democratic advantage. Clear standards for trustworthy AI will help the United States remain a center of innovation and shape technology to democratic values.

The "algorithmic audit" nevertheless remains ill-defined and inexact, whether concerning social media platforms or AI systems generally. The risk is significant that inadequate audits will obscure problems with algorithmic systems and create a permission structure around poorly designed or implemented AI. A poorly designed or executed audit is at best meaningless and at worst even excuses harms that the audits claim to mitigate. Inadequate audits or those without clear standards provide false assurance of compliance with norms and laws, "audit-washing" problematic or illegal practices. Like green-washing and ethics-washing before, the audited entity can claim credit without doing the work.

To address these risks, this paper identifies the core questions that need answering to make algorithmic audits a reliable AI accountability mechanism. The "who" of audits includes the person or organization conducting the audit, with clearly defined qualifications, conditions for data access, and guardrails for internal audits. The "what" includes the type and scope of audit, including its position within a larger sociotechnical system. The "why" covers audit objectives, whether narrow legal standards or broader ethical goals, essential for audit comparison. Finally, the "how" includes a clear articulation of audit standards, an important baseline for the development of audit certification mechanisms and to guard against audit-washing.

Algorithmic audits have the potential to transform the way technology works in the 21st century, much as financial audits transformed the way businesses operated in the 20th century. They will take different forms, either within a sector or across sectors, especially for systems which pose the highest risk. But as algorithmic audits are encoded into law or adopted voluntarily as part of corporate social responsibility, the audit industry must arrive at shared understandings and expectations of audit goals and procedures. This paper provides such an outline so that truly meaningful algorithmic audits can take their deserved place in AI governance frameworks.[1]

1 "A version of this paper will be published in the Santa Clara High Technology Law Journal (www.htlj.org), Volume 39"

Introduction

Calls for audits to expose and mitigate harms related to algorithmic decision systems are proliferating[1] and audit provisions are coming into force, notably in the EU's Digital Services Act.[2] In response to these growing concerns, nearly every research organization that deals with the ethics of AI has called for the ethical auditing of algorithms, research organizations working on technology accountability have called for ethics and/or human rights auditing of algorithms,and an artificial intelligence (AI) audit industry is rapidly developing, signified by the consulting giants KPMG and Deloitte marketing their services.[3] Algorithmic audits are a way to increase accountability for social media companies and to improve the governance of AI systems more generally. They can be elements of industry codes, prerequisites for liability immunity, or new regulatory requirements.[4] Even when not expressly prescribed, audits may be predicates for enforcing data-related consumer protection law, or what US Federal Trade Commissioner Rebecca Slaughter calls "algorithmic justice," which entails civil rights protections to "limit the dangers of algorithmic bias and require companies to be proactive in avoiding discriminatory outcomes."[5]

The desire for "audits reflects a growing sense that algorithms play an important, yet opaque, role in the decisions that shape people's life chances as well as a recognition that audits have been uniquely helpful in advancing our understanding of the concrete consequences of algorithms in the wild and in assessing their likely impacts."[6] Much as financial audits transformed the way businesses operated in the 20th century, algorithmic audits can transform the way technology works in the 21st. Stanford University's 2022 AI Audit Challenge lists the benefits of AI auditing, namely: verification, performance, and governance:

> It allows public officials or journalists to verify the statements made by companies about the efficacy of their algorithms, thereby reducing the risk of fraud and misrepresentation. It improves competition on the quality and accuracy of AI systems. It could also allow governments to establish high-level objectives without being overly prescriptive about the means to get there. Being able to detect and evaluate the potential harm caused by various algorithmic applications is crucial to the democratic governance of AI systems.[7]

At the same time, inadequate audits can obscure problems with algorithmic systems and create a permission structure around poorly designed or implemented AI. Steering audit practices and associated governance to produce meaningful accountability will be essential for "algorithmic audits" to take a deserved place in AI governance frameworks. To this end, one must confront the reality that audit discourse tends to be inexact and confusing.[8] There is no settled understanding of what an "algorithmic audit" is—not for social media platforms and not generally across AI systems. Audit talk frequently bleeds into transparency talk: transparency

1 Shea Brown, Jovana Davidovic, and Ali Hasan, "The Algorithm Audit: Scoring the Algorithms That Score Us," Big Data & Society, January 28, 2021, p. 1.

2 European Commission, Digital Services Act: Commission Welcomes Political Agreement on Rules Ensuring a Safe and Accountable Online Environment, April 23, 2022.

3 See, for example, KPMG, Achieving trustworthy AI: A Model for Trustworthy Artificial Intelligence, November 24, 2020; Deloitte, Deloitte Introduces Trustworthy AI Framework to Guide Organizations In Ethical Application Of Technology In The Age Of With, August 16, 2020.

4 See, for example, International Organization of Securities Commissions (IOSCO), The use of artificial intelligence and machine learning by market intermediaries and asset managers, June 2020; Board Of Governors of The Federal Reserve System, SR 11-7: Guidance on Model Risk Management, April 4, 2011; Laurent Dupont, Olivier Fliche, and Su Yang, Governance of Artificial Intelligence in Finance, ACPR Banque de France, June 2020.

5 Rebecca Kelly Slaughter, Janice Kopec, and Mohamad Batal, "Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission," Yale Information Society Project and Yale Journal of Law & Technology 23, August 2021, p. 56.

6 Briana Vecchione, Salon Barocas, and Karen Levy, "Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies", EAAMO '21: Equity and Access in Algorithms, Mechanisms, and Optimization, Article 19, October 2021, p. 1.

7 Marietje Schaake and Jack Clark, Stanford Launches AI Audit Challenge, Stanford HAI, July 11, 2022.

8 See Jacqui Ayling and Adriane Chapman, "Putting AI Ethics to Work: Are the Tools Fit for Purpose?" AI and Ethics 2, September 12, 2021, p. 421; Ghazi Ahamat, Madeleine Chang, and Christopher Thomas, Types of Assurance in AI and the role of Standards, Centre for Data Ethics and Innovation (CDEI) Blog, April 17, 2021.

measures open up "black box" algorithms to public scrutiny and then audits are conducted once the lid is off.[9] Legal provisions and policies referring to "audit" may have in mind a self-assessment, such as an algorithmic impact assessment, or a rigorous review conducted by independent entities with access to the relevant data.[10]

There is no settled understanding of what an "algorithmic audit" is—not for social media platforms and not generally across AI systems.

This paper poses core questions that need addressing if algorithmic audits are to become reliable AI accountability mechanisms. It breaks down audit questions into the who, what, why, and how of audits. We recognize that the definition of "algorithm" is broad, context-dependent, and distinct from the definition of AI, since not all algorithms use AI.[11] But audit provisions have as their central concern an AI process—that is, as defined by the US National Artificial Intelligence Initiative Act of 2020, "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments."[12] Therefore, we use the terms "AI" and "algorithmic" audit interchangeably without insisting on any particular definition of these terms.

In posing these questions, we do not mean to suggest that audits will look the same either within a sector or across sectors. Audits of high-risk systems, such as biometric sorting in law enforcement,[13] will be different from audits of lower-risk systems, such as office utilization detection in property management.[14] The EU's proposed AI Act distinguishes among risk categories for audit and other purposes[15] and we suspect the future of audit regulation will be strongly influenced by this approach.[16] While the substantive requirements for audits will vary with risk and context, all audit regimes will have to settle the following basic questions:

Who is conducting the audit? Self-audits, independent audits, and government audits have different features and sources of legitimacy. Moreover, the credibility of auditors will depend on their professionalism, degrees of access to data, and independence.

What is being audited? Algorithms are embedded in complex sociotechnical systems involving personnel, organizational incentive structures, and business models.[17] What an audit "sees" depends on what aspects of this complex system it looks at. The audit results will also depend on when in a system's lifecycle the audit is looking. The life of an AI system starts with the choice to deploy AI, proceeding through model development and deployment (including human interactions), and carrying through to post-deployment assessment and modification.[18] An audit can touch any or all of these moments.

Why is the audit being conducted? The objective of an audit may broadly be to confirm compliance with requirements set forth in human rights standards, sector-spe-

9 Tom Cassauwers, "Opening the black box of artificial intelligence," Horizon: The EU Research and Innovation Magazine, December 1, 2020.

10 James Guszcza et al, "Why We Need To Audit Algorithms," Harvard Business Review, November 28, 2018.

11 Kristian Lum and Rumman Chowdhury, "What Is an 'algorithm'? It Depends on Whom You Ask," MIT Technology Review, February 26, 2021.

12 William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021 (NDAA FY21), Pub. L. No. 116-283, § 5002, 134 Stat. 3388 (2021).

13 Kashmir Hill, "Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match," The New York Times, December 29, 2020.

14 Patrick Sisson, "How Data Is Changing the Way Offices Are Run," The New York Times, April 27, 2021.

15 European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 2021.

16 See Charlotte Siegmann and Markus Anderljung, The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market, Centre for the Governance of AI, August 2022.

17 Joshua A. Kroll, "Responsible AI Is a Management Problem, Not a Purchase," The Regulatory Review, July 4, 2022.

18 Jennifer Cobbe, Michelle Seng Ah Lee, and Jatinder Singh, Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems, ACM Conference on Fairness, Accountability, and Transparency (FAccT '21), March 2021, p. 599.

cific regulations, or particularized measures of fairness, non-discrimination, data protection, or to provide systemic governance and safeguard individual rights.[19] Another audit objective might be to assure stakeholders that the system functions as represented, including that the system is fair, accurate, or privacy-protecting. This is akin to the financial auditor certifying that financial statements are accurate. A subsidiary goal of either the compliance or assurance audit is to create more reflexive internal processes around the development and deployment of AI systems.[20] The audit's objectives will have significant impact on what gets audited by whom, and what sort of accountability regime the audit fits into. Consideration of an audit's purpose must all account for potential costs, financial or otherwise, for the audited entity and regulatory agencies.

How is the audit being conducted? The methodology and standards by which the audit is conducted will affect its legitimacy.[21] Common approaches generated by standard-setting bodies, codes of conduct, or other means of consensus building will also make it easier to compare audit results and act on them.

This paper first surveys the current state of algorithmic audit provisions in European and North American (often draft) law that would force greater algorithmic accountability through audit or related transparency requirements. We then identify governance gaps that might prevent audits, especially in the case of digital platform regulation, from effectively advancing the goals of accountability and harm reduction.

Algorithmic Audits: Accountability or False Assurance

Algorithmic audits can potentially address two related problems: the opacity of machine learning algorithms and the illegal or unethical performance of algorithmic systems.[22] At the same time, audits can function as window-dressing, concealing fundamental social and technical deficiencies through false assurance.

Accountability

Concern has been growing over what Frank Pasquale called in his 2016 pathbreaking book *The Black Box Society*.[23] Algorithmic processes make recommendations or decisions based on data processing and computational models that can be difficult to interrogate or understand—both within a firm and without.[24] Algorithms range in complexity from relatively simple decision trees, which are easily understood, to complex machine learning processes whose "rationales" are difficult for any human to understand. The Netherlands government in its audit framework provides the following examples of different algorithms:

- Decision trees, such as those deciding on the amount or duration of a benefit payment.
- Statistical machine learning models, such as those detecting applications with a high risk of inaccuracy to prompt additional checks.

19 Lorna McGregor, Daragh Murray, and Vivian Ng, "International Human Rights Law as a Framework For Algorithmic Accountability," International & Comparative Law Quarterly 68, 2, 2019; Margot E. Kaminski and Gianclaudio Malgieri, "Algorithmic impact assessments under the GDPR: producing multi-layered explanations," International Data Privacy Law 11, 2, April 2021.

20 See, for example, Bogdana Rakova et al, "Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices," Proceedings of the ACM on Human-Computer Interaction 5, no. 7, April 13, 2021; Jakob Mökander and Maria Axente, "Ethics-Based Auditing of Automated Decision-Making Systems," AI and Society, October 27, 2021.

21 See, for example, Adriano Koshiyama, Emre Kazin, and Philip Treleaven, Familiar methods can help to ensure trustworthy AI as the algorithm auditing industry grows, OECD AI Policy Observatory, August 10, 2021.

22 See, for example, Inioluwa Deborah Raji et al., "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing", Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, January 2020, pp. 33, 38; James Guszcza et al, "Why We Need to Audit Algorithms," Harvard Business Review, November 28, 2018.

23 Frank Pasquale, The Black Box Society: The Secret Algorithms That Control Money and Information, Harvard University Press, 2016.

24 Andrew D. Selbst and Solon Barocas, "The Intuitive Appeal of Explainable Machines," Fordham Law Review 87, no. 3 January 2018.

- Neural networks, such as facial recognition software used to detect human trafficking by examining photos on a suspect's phone.[25]

Opacity concerns are especially acute as the algorithmic process becomes more dependent on machine learning models. Particularly when they are used to inform critical determinations such as who gets hired[26] or policed,[27] the opacity of these processes can compromise public trust and accountability[28] and make it more difficult to challenge or improve decision-making.[29]

A related issue is the performance of algorithmic systems. It is well documented that machine learning algorithms can recapitulate and exacerbate existing patterns of bias and disadvantage.[30] Social media algorithms can accelerate and broaden the spread of harmful information.[31] Algorithms involved in workplace productivity[32] and educational performance[33] have been found to misjudge and therefore misallocate benefits. These problems of performance are not caused by opacity, but they are made worse when the defects are hidden in unintelligible and secret systems.

Particularly when they are used to inform critical determinations such as who gets hired or policed, the opacity of these processes can compromise public trust and accountability and make it more difficult to challenge or improve decision-making.

It is notoriously difficult to regulate technology for many reasons, including lack of institutional capacities[34] and the likelihood that technological change outpaces regulatory process.[35] Insisting on more transparency around the design and performance of algorithms is one response to the opacity problem.[36] Methods to force greater transparency include conducting algorithmic impact statements,[37] requiring researcher access to data,[38] and making aspects of government algorithmic systems transparent through records

25 Netherlands Court of Audit, Understanding Algorithms, January 26, 2021.

26 Ifeoma Ajunwa, "The Auditing Imperative for Automated Hiring," Harvard Journal of Law and Technology 34, no.2, March 24, 2021.

27 Elizabeth E. Joh, "Feeding the Machine: Policing, Crime Data, & Algorithms," William and Mary Bill of Rights Journal 26, no. 3, December 2017.

28 See Teresa M. Harrison and Luis Felipe Luna-Reyes, "Cultivating trustworthy artificial intelligence in digital government," Social Science Computer Review 40, no. 2, 2022; Cynthia Dwork and Martha Minow, "Distrust of Artificial Intelligence: Sources & Responses from Computer Science & Law," Daedalus 151, no. 2, 2022; Baobao Zhang and Allan Dafoe, "US public opinion on the governance of artificial intelligence," Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, 2020.

29 See Margot E. Kaminski and Jennifer M. Urban, "The Right to Contest AI," Columbia Law Review 121, no. 7, 2021, p. 1965.

30 Safiya Umoja Noble, Algorithms of Oppression, NYU Press, 2018, p. 24; Solon Barocas and Andrew D. Selbst, "Big Data's Disparate Impact," California Law Review 104, 2016, p. 674; Pauline T. Kim, "Data-Driven Discrimination at Work," William and Mary Law Review 58, no. 4, February 2017), p. 875.

31 See Global Internet Forum to Counter Terrorism, Content-Sharing Algorithms, Processes, and Positive Interventions Working Group Part 1: Content-Sharing Algorithms & Processes , July 2021; Florian Saurwein and Charlotte Spencer-Smith, "Automated trouble: The role of algorithmic selection in harms on social media platforms," Media and Communication 9, no. 4, 2021; Wen-Ying Sylvia Chou and Anna Gaysynsky, "A prologue to the special issue: health misinformation on social media," American Journal of Public Health, 110, suppl. 3, 2020, S270-S272.

32 Jodi Kantor and Arya Sundaram, "The Rise of the Worker Productivity Score," The New York Times, August 14, 2022.

33 Amany Elbanna and Jostein Engesmo, "A-Level Results: Why Algorithms Get Things So Wrong—and What We Can Do to Fix Them," The Conversation, August 19, 2020 .

34 See, for example, Rebecca Crootof and B.J. Ard, "Structuring Techlaw," Harvard Journal of Law and Technology 34, forthcoming, pp. 347, 376.

35 See, for example, Gary E. Marchant, "The Growing Gap Between Emerging Technologies and the Law," in Gary Marchant, Braden Allenby, and Joseph Herkert (eds.), The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem, Springer Science and Business Media B.V., 2011.

36 Robert Brauneis and Ellen P. Goodman, "Algorithmic Transparency for the Smart City," Yale Journal of Law and Technology 20, 2018, p. 129.

37 Andrew D. Selbst, "An Institutional View of Algorithmic Impact Assessments," Harvard Journal of Law and Technology 35, no. 1, 2021.

38 Nathaniel Persily, "A Proposal for Researcher Access to Platform Data: The Platform Transparency and Accountability Act," Journal of Online Trust and Safety, October 2021, p. 2.

requests.[39] It must be recognized, however, that transparency alone is of limited utility for complex algorithmic systems.[40] Commonly used AI models make predictions based on classifications that an algorithm has "learned." For example, an algorithm might "learn" from old data to classify what is a high-risk loan or a desirable employee.[41] The model will then use these learnings to make predictions about new scenarios.[42] How the model converts learnings into predictions is not easy to render transparent.[43] The mere production of computer code or model features will be insufficient to make transparency meaningful.[44] The goal of making an algorithm legible to humans is now often expressed in terms of explainability[45] or interpretability,[46] rather than transparency. To this end, computer scientists are working in partnership with others to create "explainable AI" or xAI.[47] Yet so far at least, aspirational explainability cannot be relied upon either for effective communication about how algorithmic systems works or for holding them to account.[48]

If well-designed and implemented, audits can abet transparency and explainability.[49] They can make visible aspects of system construction and operation that would otherwise be hidden. Audits can also substitute for transparency and explainability. Instead of relying on those who develop and deploy algorithmic systems to explain or disclose, auditors investigate the systems themselves.[50] This investigation can address the black box problem by providing assurance that the algorithm is working the way it is supposed to (for example, accurately) and/or that it is compliant with applicable standards (for example, non-discrimination). To the extent that there are problems, the audit will ideally turn them up and permit redress and improvement. Poor audit design and implementation will hinder the delivery of these benefits and actually do harm.

False Assurance

Experience with audits in other contexts raises the specter of false assurance. A firm that has audited itself or submitted to inadequate auditing can provide false assurance that it is complying with norms and laws, possibly "audit-washing" problematic or illegal practices. A poorly designed or executed audit is at best meaningless. At worst, it can deflect attention from or even excuse harms that the audits are supposed to mitigate.[51] Audit washing is a cousin of "green washing" and

39 Hannah Bloch-Wehba, "Access to Algorithms", Fordham Law Review 88, August 2020.

40 See Joshua A. Kroll et al, "Accountable Algorithms," University of Pennsylvania Law Review 165, 2017, pp. 657-660

41 Gabriel Nicholas, "Explaining Algorithmic Decisions," Georgetown Law Tech Review 4, 2020, p 714.

42 Brauneis and Goodman, "Algorithmic Transparency for the Smart City," pp. 113-114

43 See, for example, Katherine J. Strandburg, "Rulemaking and Inscrutable Automated Decision Tools," Columbia Law Review 119, no. 7, 2019, p. 1862; David Freeman Engstrom and Daniel E. Ho, "Algorithmic Accountability in the Administrative State," Yale Journal on Regulation 37, 2020, p. 821; Brent Mittelstadt, Chris Russell, and Sandra Wachter, Explaining Explanations in AI, in FAT* '19: Conference on Fairness, Accountability, and Transparency, 2019; Mike Ananny and Kate Crawford, "Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability," New Media & Society 20, no. 3, March 2018, p. 981.

44 See Maayan Perel and Niva Elkin-Koren, "Black box tinkering: Beyond disclosure in algorithmic enforcement," Florida Law Review 69, 2017, p. 181; Cansu Safak and Imogen Parker, Meaningful transparency and (in) visible algorithms, Ada Lovelace Institute, October 15, 2020; Matthew Gooding, "Elon Musk's plan for an open-source algorithm won't solve Twitter's problems," techmonitor.ai, April 26, 2022.

45 Ashley Deeks, "The Judicial Demand for Explainable Artificial Intelligence," Columbia Law Review; Engstrom and Ho, "Algorithmic Accountability in the Administrative State," p. 804.

46 Cynthia Rudin, "Stop Explaining Black Box Models for High Stakes Decisions and Use Interpretable Models Instead," Nature Machine Intelligence 1, p. 2018.

47 P. Jonathan Phillips et al, Four Principles of Explainable Artificial Intelligence, National Institute of Science and Technology (NIST), August 2020; David Gunning et al, "DARPA's Explainable AI (XAI) Program: A Retrospective." Applied AI Letters 2, no. 4, 2021, p. e61.

48 For a critique of xAI, see, for example, Nicholas, "Explaining Algorithmic Decisions."

49 Pauline T. Kim, "Auditing Algorithms for Discrimination," University of Pennsylvania Law Review Online 166, 2017, p. 190.

50 Danielle Keats Citron, "Technological Due Process," Washington University Law Review 85, 2008, pp. 1249, 1305; Kate Crawford and Jason Schultz, "Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms," Boston College Law Review 55, no. 1, 2014, pp. 121-124.

51 Julian Jaursch, "Why The EU Needs To Get Audits For Tech Companies Right," Techdirt, August 19, 2021.

Case Study: Meta's Civil Rights Audit

The example of Meta's civil rights audit in 2020 illustrates the limitations of self-audits and second-party audits, especially without any accountability mechanism to ensure that audited firms implement changes in response to audit findings.

Following pressure from both the US Congress and civil rights groups, in 2018 Facebook (now Meta) commissioned a civil rights audit led by Laura Murphy, a former American Civil Liberties Union official, and Megan Cacace, a partner at Relman Colfax. They released a series of reports culminating in an 89-page audit report in July 2020.[1]

The report generated inflammatory headlines highlighting the audit's damning findings. Most notably, the auditors found that Facebook's decision to keep up certain posts from President Donald Trump represented "significant setbacks for civil rights." They criticized Facebook's response to hate speech and misinformation on the platform, stating, "Facebook has made policy and enforcement choices that leave our election exposed to interference by the President and others who seek to use misinformation to sow confusion and suppress voting."[2] The audit also addressed key issues where Facebook's policies around labelling, takedowns, and its advertising library were found lacking, including on COVID-19, election misinformation, and extremist or white-nationalist content. The audit acknowledged Facebook's stated commitments to civil rights—including policies undertaken to combat voter suppression and hiring a senior official for civil rights advancement—but expressed concern that other decisions undermined progress. The auditors concluded:

> Unfortunately, in our view Facebook's approach to civil rights remains too reactive and piecemeal. Many in the civil rights community have become disheartened, frustrated and angry after years of engagement where they implored the company to do more to advance equality and fight discrimination, while also safeguarding free expression.[3]

While scathing in its indictment of Facebook's policies, the report was nevertheless greeted with a certain degree of skepticism by the civil rights groups that had pushed for its commissioning, as it notably contained no concrete commitments or guarantees from Facebook of future policy changes. Rashad Robinson, president of Color of Change, told National Public Radio that "The recommendations coming out of the audit are as good as the action that Facebook ends up taking. Otherwise, it is a road map without a vehicle and without the resources to move, and that is not useful for any of us."[4]

The audit's proposed solutions—even if enacted—also seemed to mirror many of Facebook's own proposals proffered under criticism. As tech journalist Casey Newton wrote at The Verge,

> The auditors' view of Facebook is one in which the company looks more or less the same as it does today, except with an extra person in every meeting saying "civil rights." That would surely do some good. But it would not make Facebook's decisions any less conse-

1 Facebook's Civil Rights Audit, Facebook's Civil Rights Audit Report—Final Report, July 8, 2020.

2 Ibid., p. 10.

3 Ibid., p. 8.

4 Shannon Bond, "Report Slams Facebook for 'Vexing and Heartbreaking Decisions' on Free Speech," National Public Radio, July 8, 2020.

quential, or reduce the chance that a future content moderation decision or product problem stirs up the present level of outrage. The company could implement all of the auditors' suggestions and nearly every dilemma would still come down to the decision of one person overseeing the communications of 1.73 billion people each day.[5]

The report also focused solely on the United States, at a time when Facebook's human rights record in non-US and non-Anglophone countries was undergoing substantial scrutiny. A human rights impact assessment commissioned in India was strongly criticized by human rights groups, who accused Facebook executives of delaying and narrowing the report.[6]

While Facebook clearly "failed" its civil rights audit, the meaning of failure must be questioned when the resulting recommendations were toothless.

While Facebook clearly "failed" its civil rights audit, the meaning of failure must be questioned when the resulting recommendations were toothless. Chief Operating Officer Sherly Sandberg responded to the report in a blog post where she described the findings as "the beginning of the journey, not the end" and promised to "put more of their [auditors] proposals into practice," but that Facebook would not make "every change they call for."[7] Can an audit be considered a success if the most concrete outcome is a vague promise to consider or test a new policy?

The revelations by whistleblower Frances Haugen in the fall of 2021 renewed criticism of the same shortcomings underscored by the audit, highlighting the lack of progress made since its publication. Auditor Laura Murphy, in a 2021 report on guidelines for such audits, wrote that "Facebook's recent crisis has alienated some key stakeholders and overshadowed many of the important and groundbreaking tangible outcomes yielded by its civil rights audit," echoing the audit's previous criticism of the one-step-forward, two-steps-back nature of the problem and the platform's response.[8]

Civil rights audits have become a common response to criticism, undertaken across industries and including tech giants like Google, Microsoft, Amazon, and Uber. But these remain voluntary and when undertaken lack transparency or common metrics and standards. The who of this audit was clear, but the what and how did not conform to any predetermined standards or frameworks. The why was also unclear, because despite the audit's findings of Facebook's shortcomings, there was no mechanism or benchmark to enforce change. The definition of success or failure is arbitrary, and enforcement or consequences are lacking. Reputational damage is insufficient to force needed reforms, echoing criticisms also lodged against Facebook's Oversight Board or voluntary obligations like the Global Network Initiative. While the auditors demonstrated necessary independence and delivered a critical report, the risk of audit-washing remains without broader standards and methodology to reliably replicate and compare audits. Facebook's civil rights audit—while not explicitly related to algorithms—illustrates the limits of auditing without clear guidelines and accountability mechanisms.

---

5 Casey Newton and Zoe Schiffer, "What a damning civil rights audit missed about Facebook," The Verge, July 10, 2020.

6 Newley Purnell, "Facebook is Stifling Independent Report on its Impact in India, Human Rights Groups Say," The Wall Street Journal, November 12, 2021.

7 Sheryl Sandberg, Making Progress on Civil Rights—But Still a Long Way to Go, Meta, July 8, 2020.

8 Laura W. Murphy, The Rationale For and Key Elements of a Business Civil Rights Audit, The Leadership Conference on Civil and Human Rights, 2021.

"ethics washing"—the acquisition of sustainability or ethical credibility through cosmetic or trivial steps.[52]

One common way for audits to fall into audit washing is when a firm self-audits without clear standards. For example, Meta conducted a human rights impact assessment of its own company's (Facebook's) role in inciting the 2018 genocide in Myanmar. The review "was considered a failure that acted more like 'ethics washing' than anything substantive."[53] Another common pitfall in the technology space is for a firm to profess adherence to human rights standards without actually designing its systems to deliver on them.[54]

Even when outside checks are ostensibly in place, systems of assurance may simply mask wrongdoing. The US Federal Trade Commission (FTC) will often enter into settlement agreements with companies for privacy violations and, as part of the agreement, require companies to obtain an outside assessment of the firm's privacy and security program.[55] An assessment is a less rigorous form of review than audit because it looks at conformity with the firm's own goals as opposed to conformity with third-party standards. Chris Hoofnagle has shown that success in these privacy assessments bears little relation to actually successful privacy practices. For example, Google submitted a privacy assessment suggesting perfect compliance even though, "during the assessment period, Google had several adverse court rulings on its services, including cases ... suggest[ing] the company had violated federal wiretapping laws."[56]

Algorithmic Audits in Legislation and Governmental Inquiries

Legislation, proposed or enacted, around the world would promote or require algorithmic audits, especially for large online platforms. The following reviews an assortment of leading algorithmic audit legislation in the EU, the United Kingdom, the United States, and individual US states.

The European Union

The EU's landmark Digital Services Act (DSA) requires in Articles 26 and 27 that very large online platforms (VLOPs) conduct annual systemic risk assessments of online harms and take appropriate mitigating measures.[57] The DSA also requires VLOPs that use recommendation systems to reveal in their Terms of Service the primary parameters used by algorithmic amplification systems.[58] Article 28 of the DSA requires VLOPs to submit yearly to external audits to certify that they have complied with these risk mitigation and reporting requirements, but it does not mandate that the auditors actually conduct an independent risk assessment. Earlier DSA drafts were criticized for not requiring sufficient independence for auditors.[59] The final version provides some detail about auditor independence.[60] It remains the case, however, that the task of auditors is merely to "verify that the VLOP has complied with the obligation to perform a risk assessment and that the mitigation measures identified by the VLOP are coherent with its own findings about the systemic risks posed by its own services."[61] Finally, the DSA proposes a mechanism in Article 31 for facilitating data access to vetted researchers and others, in part so they can explore algorithmic systems such as recommender

52 See, for example, Elettra Bietti, "From Ethics Washing to Ethics Bashing: A Moral Philosophy View on Tech Ethics," Journal of Social Computing 2, no. 3, September 2021.

53 Mark Latonero and Aaina Agrawal, "Human Rights Impact Assessments For AI: Learning From Facebook's Failure In Myanmar," Carr Center for Human Rights Policy, March 19, 2021.

54 Karen Yeung, Andrew Howes, and Ganna Pogrebna, "AI Governance by Human Rights Centred-Design, Deliberation and Oversight: An End to Ethics Washing," Markus D. Dubber, Frank Pasquale, and Sunit Das (eds.), in The Oxford Handbook of Ethics of AI, Oxford University Press, 2020.

55 Chris Jay Hoofnagle, "Assessing the Federal Trade Commission's Privacy Assessments," IEEE Security & Privacy 14, no. 2, 2016.

56 Ibid., p. 62.

57 Luca Bertuzzi, "EU Institutions Reach Agreement on Digital Services Act," EURACTIV, April 23, 2022.

58 James Vincent, "Google, Meta, and Others Will Have to Explain Their Algorithms under New EU Legislation," The Verge, April 23, 2022.

59 Ilaria Buri and Joris van Hoboken, The Digital Services Act (DSA) proposal: a critical overview, DSA Observatory, October 28, 2021.

60 European Commission Amendments adopted by the European Parliament on 20 January 2022 on the proposal for a regulation of the European Parliament and of the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC, updated May 13, 2022.

61 Buri and van Hoboken, "Digital Services Act (DSA) proposal," p. 37.

systems.[62] In this way, principally academic researchers are expected to perform an auditing function, although the scope and definition of vetted researcher access has yet to be defined. Non-EU academics, researchers, and civil society groups also hope to be able to benefit from some of these transparency requirements.

Other EU laws or initiatives that are part of the algorithmic audit and transparency ecosystem include the Platform-to-Business Regulation and the New Deal for Consumers, which mandate disclosure of the general parameters for algorithmic ranking systems to business users and consumers respectively.[63] The General Data Protection Regulation (GDPR) sets rules for the profiling of individuals and related automated decision-making and gives users the "right to explanation" about algorithmic processes.[64] Margot Kaminski observes that GDPR guidelines contemplate at least internal audits of algorithms "to prevent errors, inaccuracies, and discrimination on the basis of sensitive ... data" in individual automated decision-making.[65] Commentators predict that this right, as well as entitlements to access collected data, will lead to robust independent audits.[66] The EU's Digital Markets Act in Article 13 obliges designated gatekeepers to submit their techniques of data-profiling consumers to an independent audit, but it does not specify procedures for the audit.[67]

The EU's draft Artificial Intelligence Act proposes a risk-based approach to AI regulation along a sliding scale of potential harms, and it requires in Article 61 that providers of high-risk AI systems conduct "conformity assessments" before their products enter the European market.[68] This is an internal audit to ensure that governance of the AI is compliant with regulation. The Act would also create a post-market monitoring requirement for high-risk AI systems. Very high-risk AI systems defined as those intended for use in real-time or remote biometric identification may require external audits.[69] This approach to high-risk AI systems involves a combination of self-regulation, voluntary adherence to standards, and government oversight.[70]

The United States

In the United States, a 2016 report by the Obama administration on algorithms and civil rights encouraged auditing.[71] In Congress, the Algorithmic Accountability Act was re-introduced in 2022 and would require the FTC to create regulations and structures for companies to carry out assessments and provide transparency around the impact of automated decision-making.[72] Covered entities would be required to "perform ongoing evaluation of any differential performance associated with data subjects' race, color, sex, gender, age, disability, religion, family-, socioeconomic-, or veteran status." This seems like a step towards greater algorithmic fairness but raises the question of what kind of fairness counts and how should it be measured. Scholars have pointed out that there are many ways to measure "differential performance" and definitions of fairness differ within and between disciplines of law, computer science, and

62 Paddy Leerssen, "Platform Research Access in Article 31 of the Digital Services Act: Sword without a Shield?", Verfassungsblog, September 7, 2021.

63 European Commission, Platform-to-Business Trading Practices, June 7, 2022; Věra Jourová, The New Deal for Consumers: What Benefits Will I Get as a Consumer?, European Commission, November 2019.

64 GDPR Article 22.

65 Margot E. Kaminski, "The Right to Explanation, Explained," Berkeley Technology Law Journal 34, 2019, p. 206.

66 Casey, Farhangi, and Vogl, "Rethinking Explainable Machines," pp. 150-151.

67 European Commission, Proposal For a Regulation of The European Parliament And Of The Council On Contestable And Fair Markets In The Digital Sector (Digital Markets Act), EUR-Lex, December 15, 2020.

68 European Commission, Regulation of the European Parliament and of the Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) And Amending Certain Union Legislative Act, EUR-Lex, April 21, 2021.

69 Natasha Lomas, "Europe's AI Act contains powers to order AI models destroyed or retrained, says legal expert," TechCrunch, April 1, 2022.

70 Margot E. Kaminski, "Regulating the Risks of AI," Boston University Law Review 103, forthcoming, posted August 19, 2022, pp. 51-54.

71 Executive Office of the President, Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights, May 2016.

72 US Congress, Senate, Algorithmic Accountability Act of 2022, S.3572, 117th Congress, 2nd sess., introduced in the Senate February 3, 2022.

others.[73] Moreover, fairness may conflict with other desirable goals of accuracy, efficiency, and privacy.[74]

The Digital Services Oversight and Safety Act, introduced in 2022, would require the FTC to create regulations for large online platforms, requiring them to assess "systemic risks" (including the spread of illegal content and goods and violation of community standards with an "actual or foreseeable negative effect on the protection of public health, minors, civic discourse, electoral processes, public security, or the safety of vulnerable and marginalized communities").[75] The platforms would be required to commission an annual independent audit of their risk assessments and submit these to the FTC. The American Data Privacy and Protection Act, released as a discussion draft in 2022, would require data processors that knowingly develop algorithms to collect, process, or transfer covered data to evaluate algorithmic design (preferably through an independent audit), including any training data used to develop the algorithm, to reduce the risk of civil rights harms.[76]

The White House released a Blueprint for an AI Bill of Rights in October 2022 that explicitly mentions auditing.

Other proposed legislation for online platforms would require transparency that might, ultimately, foster the development of independent platform audits. The Algorithmic Justice and Online Platform Transparency Act would prohibit discriminatory use of personal information in algorithmic processes and require transparency in algorithmic decision-making.[77] The Social Media NUDGE Act would require researcher and government study of algorithms and platform cooperation in reducing the spread of harmful content, with oversight by the FTC.[78]

The National Institute of Standards and Technology (NIST) published a draft risk management framework for AI systems in March 2022 in which it recommends the evaluation of such systems by an "independent third party or by experts who did not serve as front-line developers for the system, and who consults experts, stakeholders, and impacted communities."[79] The NIST framework will ultimately be a guiding set of principles,[80] not binding legislation, and avoids setting explicit risk thresholds for companies.

US state-level lawmakers have introduced legislation requiring algorithmic auditing for civil rights in certain contexts. New York City published an AI strategy and a new law coming into force in January 2023 will require entities using AI-based hiring tools to commission independent bias audits and disclose to applicants how AI was used, with fines for using undisclosed or biased systems.[81] In the limited context of pretrial risk assessment tools, the state of Idaho requires algorithmic transparency and open access to the public for "inspection, auditing, and testing" of those tools.[82] Washington D.C.'s Attorney General has proposed a bill prohibiting algorithmic discrimination with respect to eligibility for "important life opportunities", and

73 Richard N. Landers and Tara S. Behrend, "Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models," American Psychologist, February 14, 2022, pp. 2-3.

74 Jess Whittlestone et al, "The Role and Limits of Principles in AI Ethics: Towards A Focus on Tensions," AIES 2019 Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, January 2019.

75 US Congress, House of Representatives, Digital Services Oversight and Safety Act of 2022, H.R.6796, 117 Congress, 2nd sess., introduced in the House February 18, 2022.

76 US Congress, House of Representatives, American Data Privacy and Protection Act, H.R.8152, 117 Congress, 2nd sess., introduced in the House June 21, 2022.

77 US Congress, Senate, Algorithmic Justice and Online Platform Transparency Act, S.1896, 117th Congress, 1st sess., introduced in the Senate May 27, 2022.

78 US Congress, Senate, Social Media NUDGE Act, S.3608, 117th Congress, 2nd sess., introduced in the Senate February 9, 2022.

79 National Institute of Standards and Technology (NIST), AI Risk Management Framework: Initial Draft, March 17, 2022.

80 David Matthews, "How the US plans to manage artificial intelligence," Science|Business, May 19, 2022.

81 The Mayor's Office of the Chief Technology Officer (NYC CTO), AI Strategy: The New York City Artificial Intelligence Strategy, October 2021; The New York City Council, Automated Employment Decision Tools, Int 1894-2020, enacted December 11, 2021; The New York City Council, Int. No. 1894-A, amended November 1, 2021.

82 Justia Law, 2020 Idaho Code: Title 19 - Criminal Procedure: Chapter 19 - Mode of Trial - Formation of Trial Jury - Postponement of Trial, accessed July 23, 2022.

would require entities to audit their decisions and retain a five-year audit trail.[83]

Finally, the White House released a Blueprint for an AI Bill of Rights in October 2022 that explicitly mentions auditing. Automated systems "should be designed to allow for independent evaluation" including by third-party auditors, and with attendant mechanisms in place to ensure speed, trustworthy data access, and protections to ensure independence. It also prescribes independent audits to ensure "accurate, timely, and complete data."[84] These non-binding principles are meant to "lay down a marker for the protections that everyone in America should be entitled to" and as a "beacon" for the "whole of government," according to Alondra Nelson, deputy director for science and society at the Office of Science and Technology Policy at the Office of Science and Technology Policy, in an interview with the *Washington Post* following its release.[85]

Canada

The Canadian government's Algorithmic Impact Assessment Tool and the Directive on Automated Decision-Making work in tandem and are designed to apply across a range of automated decision-making systems.[86] The Algorithmic Impact Assessment Tool questionnaire is a scorecard used to determine the impact level of an automated decision system. The directive imposes requirements regardless of impact level, including requirements for licensed software, transparency of government-owned code, bias testing, data quality and security assessment, legal consultations, redress for clients, and effectiveness reporting.[87] Additional requirements are also imposed according to the impact level, which can include peer review, transparency, human intervention, contingency measures, or employee training. Algorithmic impact assessments are mandatory for federal government institutions, with the exception of the Canada Revenue Agency.[88] The Expert Group on Online Safety, which convened to provide consultation on the Canadian Online Safety Bill, recommended in its final report a risk-based approach with ex ante and ex post elements, in which a digital safety commissioner would have the power to conduct audits, backed by strong enforcement powers.[89]

Australia

The 2021 News Media Bargaining Code governs commercial relationships between Australian news businesses and digital platforms.[90] It requires designated platforms to pay local news publishers for content linked on their platform and also requiring notice for changes to platform algorithms.[91] Proposed amendments to the bargaining code would empower the Australian Competition and Consumer Commission (ACCC) to conduct regular audits of the digital platform's algorithms and automated decision systems, thereby creating a formal third-party monitoring role with the code. The proposal reads: "Designated digital platforms would be required to provide the ACCC with full access to information about relevant algorithms and automated decision systems as the Commission may require to assess their impact on access to Australian news media content."[92]

83 Council of the District of Columbia, Stop Discrimination by Algorithms Act of 2021 , B24-558; Martin Austermuhle, "D.C. attorney general introduces bill to ban 'algorithmic discrimination'," National Public Radio, December 10, 2021.

84 The White House, Blueprint for an AI Bill of Rights, October 2022.

85 Cristiano Lima, "White House unveils 'AI Bill of Rights' as 'call to action' to rein in tool," The Washington Post, October 4, 2022.

86 Treasury Board of Canada Secretariat, Algorithmic Impact Assessment tool; Treasury Board of Canada Secretariat, Directive on Automated Decision-Making, Government of Canada, modified April 1, 2021.

87 Christine Ing, Michael Scherman, and Drew Wong, Federal Government's Directive on Automated Decision-Making: Considerations and Recommendations, McCarthy Tetrault LLP, April 13, 2019.

88 Benoit Deshaies and Dawn Hall, Responsible use of automated decision systems in federal government, Statistics Canada, December 1, 2021.

89 Government of Canada, Summary of Session Four: Regulatory Powers, modified May 13, 2022; Government of Canada, Concluding Workshop Summary, modified July 8, 2022.

90 Australian Competition and Consumer Commission (ACCC), News Media Bargaining Code, accessed July 23, 2022.

91 Asha Barbaschow, "Media Bargaining Code amendments include a more 'streamlined' algorithm change notice," ZDNet, July 12, 2022.

92 Parliament of Australia, Call For Tech Giants to Face Regular ACCC Algorithm Audits, January 22, 2021.

The United Kingdom

The draft UK Online Safety Bill gives regulator Ofcom significant investigatory power over platforms,[93] including the ability to audit algorithms of regulated entities.[94] Those entities must conduct risk assessments and then take steps to mitigate and manage identified risks of particular types of illegal and harmful content. Some service providers will also be required to publish transparency reports. The Information Commissioner's Office has developed draft guidance on an AI Auditing Framework for technologists and compliance officers focused on the data protection aspects of building AI systems.[95] In addition, the Centre for Data Ethics and Innovation (CDEI), which is part of the Department for Digital, Culture, Media and Sport, has provided a Roadmap to an Effective AI Assurance Ecosystem.[96] While not focused on AI audits, the CDEI roadmap lays out a range of audit and audit-like steps that help to create AI "assurance." The terms impact assessment, audit, and conformity assessment all show up in EU and UK legal instruments with particular meanings that are not the same as CDEI's.

Algorithmic Auditing Provision Holes

The above survey of algorithmic audit provisions illustrates how accountability mechanisms aimed at mitigating harms from online platforms are nested in broader AI governance structures. As algorithmic audits are encoded into law or adopted voluntarily as part of corporate social responsibility, it will be important for the audit industry to arrive at shared understandings and expectations of audit goals and procedures, as happened with financial auditors. The algorithmic audit industry will have to monitor compliance not only of social media algorithms, but also of hiring, housing, health care, and other deployments of AI systems. AI evaluation companies are receiving significant venture capital funding and are certifying algorithmic processes.[97] Still, according to Twitter's Rumman Chowdhury, the field of reputable auditing firms is small—only 10 to 20.[98] Audits will not advance trustworthy AI or platform accountability unless they are trustworthy themselves. The following sets out basic questions that need to be addressed for algorithmic audits to be a meaningful part of AI governance.

Who: Auditors

Inioluwa Deborah Raji, a leading scholar of algorithmic audits, argues that the audit process should be interdisciplinary and multistaged as it plays out, both internally for entities developing and deploying AI systems and externally for independent reviewers of those systems.[99]

Internal auditors, also known as first-party auditors, can intervene at any stage of the process. Such auditors have full access to the system components before deployment and so are able to influence outcomes before the fact. The auditing entity's goals influence the scope of the internal audit, which can focus on a technical overview, ethical considerations and harm prevention goals, or strictly legal compliance. An internal audit cannot alone give rise to public accountability and could be used to provide unverifiable assertions that the AI has passed legal or ethical standards. The proposed Algorithmic Accountability Act in the United States seems to call for first-party audits that a company will conduct on its own.[100] The same is true of the audit provisions in the GDPR. The Federal Reserve and Office of the Comptroller of the Currency's SR 11-7 guidance on model risk management suggests that an internal auditing team be different from

93 United Kingdom Government, Draft Online Safety Bill, May 2021.

94 United Kingdom Government, Findings from the DRCF Algorithmic Processing Workstream—Spring 2022, April 28, 2022.

95 Information Commissioner's Office, Guidance on the AI auditing framework: Draft guidance for consultation, accessed July 23, 2022.

96 United Kingdom Government, The roadmap to an effective AI assurance ecosystem, December 8, 2021.

97 Kate Kaye, "A new wave of AI auditing startups wants to prove responsibility can be profitable," Protocol, January 3, 2022.

98 Alfred Ng, "Can Auditing Eliminate Bias from Algorithms?", The Markup, February 23, 2021.

99 Raji et al., "Closing the AI accountability gap"; Inioluwa Deborah Raji et al, "Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance," AIES '22, Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, June 9, 2022.

100 US Congress, House of Representatives, Algorithmic Accountability Act of 2022, H.R.6580, 117 Congress, 2nd sess., introduced in the House February 3, 2022.

the team developing or using the tool subject to audit.[101] A number of commentators have called for increased rigor around internal auditing. Ifeoma Ajunwa, for example, proposes mandatory internal and external auditing for hiring algorithms.[102] Shlomit Yanisky-Ravid and Sean K. Hallisey propose a governmental or private "auditing and certification regime that will encourage transparency, and help developers and individuals learn about the potential threats of AI, discrimination, and the continued weakening of societal expectations of privacy."[103]

External audits necessarily look backward and will typically exhibit a range of independence from the deploying entity. The primary purpose of these audits is to signal trustworthiness and compliance to external audiences. An entity may contract with an auditor to produce a report, which is known as a second-party audit, or the auditor may come entirely from the outside to conduct a third-party audit.[104] The DSA notably calls for third-party audits and takes the first steps towards defining "independence" for third-party auditors. Yet there are no clear or agreed standards for these algorithmic auditing firms. This creates a risk of "audit-washing," whereby an entity touts that it has been independently audited when those audits are not entirely arms-length or are otherwise inadequate.[105] For example, the company HireVue marketed its AI employment product as having passed a second-party civil rights audit, only for the independence of the auditors and the scope of the audit to be drawn into question.[106]

In order to ensure a degree of consistent rigor among auditors, Ben Wagner and co-authors have called for "auditing intermediaries."[107] They recommend independent intermediaries as an alternative to government involvement in audits, as exists currently in Germany with respect to social media auditing required by the Network Enforcement Act (NetzDG). In that case, a government-affiliated entity audits the data the platforms are required to disclose about content moderation decisions.[108] Wagner and co-authors argue that auditing intermediaries, independent from both government and audited entities, can provide protection from government overreach, consistency for audited entities faced with multiple audit requirements across jurisdictions, rigor for audit consumers, and safety for personal data because of the special protections they can deploy.[109]

The history of financial auditing and the accretion of professional standards over time is instructive for how auditors can maintain independence. Financial audits were first required in England in the mid-19th century to protect shareholders from the improper actions of company directors.[110] At first, "there was no organized profession of accountants or auditors, no uniform auditing standards or rules, and no established training or other qualifications for auditors, and they had no professional status."[111]

According to John Carey's history of US accounting practices, it was not until the turn of the 20th century that financial accountants started to organize and regulate themselves as a profession.[112] It took until the 1930s for independent auditing to become institutionalized in the financial markets. What catalyzed the regimentation and ubiquity of financial audits was the federal legislation that

101 Comments from Andrew Burt and Solon Barocas in October 26, 2022 GMF workshop. See SR-11 Guidance on Model Risk Management, Federal Reserve, April 4, 2011; Comptroller's Handbook, "Model Risk Management," August 2021.

102 Ifeoma Ajunwa, "An Auditing Imperative for Automated Hiring Systems," Harvard Journal of Law and Technology, August 19, 2021.

103 Shlomit Yanisky-Ravid and Sean K. Hallisey, "Equality and Privacy by Design," Fordham Urban Law Journal, April 2019.

104 Raji et al, "Outsider Oversight."

105 Mona Sloane, "The Algorithmic Auditing Trap," OneZero, March 17, 2021.

106 Alex C. Engler, "Independent auditors are struggling to hold AI companies accountable," Fast Company, January 26, 2021.

107 Ben Wagner and Lubos Kuklis, "Establishing Auditing Intermediaries to Verify Platform Data," in Martin Moore and Damian Tambini (eds.), Regulating Big Tech: Policy Responses to Digital Dominance, Oxford University Press, 2021.

108 See, for example, Ben Wagner et al. 2020, "Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act," Barcelona, Spain: ACM Conference on Fairness Accountability and Transparency (FAT*), January 2020.

109 Wagner and Kuklis, 2021. See also Ben Wagner et al, "The next step towards auditing intermediaries," Verfassungsblog, February 23, 2022.

110 Howard B. Levy, "History of the Auditing World, Part 1," The CPA Journal, November 2020.

111 Ibid.

112 John L. Carey, "Rise of the accounting profession, v. 1. From technician to professional, 1896-1936," Guides, Handbooks and Manuals 30, 1969.

followed the stock market crash of 1929: the Securities Act of 1933 and Securities Exchange Act of 1934, which together required audited financial statements for public companies. Later interventions augmented audit oversight after the Enron financial scandal with the 2002 Sarbanes-Oxley Act[113]—which created a private nonprofit corporation to oversee audit procedures—and after the 2008 market crash with the 2010 Dodd-Frank Act[114]—which added to the requirements for independent audits and corporate audit committees, along with strengthening whistleblower protections.[115]

The legal regime surrounding audits and auditors will influence who conducts audits and with what rigor. External audits will likely require access to information that is either proprietary or otherwise closely held by the audited entity. Jenna Burrell has examined how firms invoke trade secrets to limit access to the data or code that may be necessary for audits, especially of complex machine learning systems whose training data is important to examine in an audit.[116] Even platforms that say they are interested in transparency, such as Reddit with its commitment to the Santa Clara Principles, seek to maintain secrecy to prevent adversarial actors from reverse-engineering the system.[117] External auditors will have to gain access to information in order to conduct reasonably competent inquiries. They will then have to ensure that release of relevant data is not blocked by nondisclosure agreements—these contracts between firms and audit companies could hinder the sharing necessary to compare audit results across firms and warrant public trust. Even the audit result in the controversial HireVue case can only be accessed on their website after signing a nondisclosure agreement.[118]

For internal and external auditors, the risk of legal liability will shape how the audit is conducted, ideally leading to appropriate care, but possibly leading to excessive caution. One of the hallmarks of financial audits is that independent auditors are subject to legal liability to third parties and regulators for failure to identify misstatements or knowingly abetting fraud.[119] In the algorithmic audit context, unless auditors are clear on the standards and goals of the audit, fear of liability could render their services useless. External audits conducted by researchers and journalists also come with legal risk, for example via the US Computer Fraud and Abuse Act if audited data is obtained without consent.[120] Scholars and public interest advocates raising this concern recently won a victory in the case of *Sandvig v. Barr*, where a federal judge ruled that the law "does not criminalize mere terms-of-service violations on consumer websites," and that research plans involving such violations in order to access data for study purposes could therefore go forward.[121] More protections for adversarial audits carried out by researchers or journalists without a company's consent may be required. For internal audits, rigorous examinations can turn up findings that potentially expose firms to legal liability. Erwan Le Merrer and co-authors call for a structural overhaul to create legal certainty that hold firms harmless for internal audits.[122]

113 US Congress, House of Representatives, Sarbanes-Oxley Act of 2002, H.R.3763, 107th Congress, introduced in the House February 14, 2002.

114 US Congress, House of Representatives, Dodd-Frank Wall Street Reform and Consumer Protection Act, H.R.4173, 111th Congress, introduced in the House December 2, 2009.

115 Sarah J. Williams, "The Alchemy of Effective Auditor Regulation," Lewis & Clark Law Review, July 15, 2022.

116 Jenna Burrell, "How the machine 'thinks': Understanding opacity in machine learning algorithms," Big Data & Society, June 1, 2016.

117 Prerna Juneja, Deepika Rama Subramanian, and Tanushree Mitra, "Through the Looking Glass: Study of Transparency in Reddit's Moderation Practices," Proceedings of the ACM on Human-Computer Interaction 4, no. 17, January 2019; Santa Clara Principles, "Santa Clara Principles: On Transparency and Accountability in Content Moderation," accessed July 24, 2022.

118 Hilke Schellmann, "Auditors are testing hiring algorithms for bias, but there's no easy fix," MIT Technology Review, February 11, 2021.

119 See, for example, Janne Chung et al. "Auditor liability to third parties after Sarbanes-Oxley: An international comparison of regulatory and legal reforms," Journal of International Accounting, Auditing and Taxation, January 2010; Alan Reinstein, Carl J. Pacini, and Brian Patrick Green, "Examining the current legal environment facing the public accounting profession: Recommendations for a consistent U.S. policy," Journal of Accounting, Auditing & Finance, January 9, 2017.

120 18 US Code § 1030, Fraud And Related Activity In Connection With Computers, Legal Information Institute, undated.

121 Sandvig v. Barr, 451 F. Supp. 3d 73, 76 (D.D.C. 2020).

122 Erwan Le Merrer, Ronan Pons, and Gilles Trédan, "Algorithmic audits of algorithms, and the law," HAL Open Science, February 22, 2022.

What/When: What Is Actually Being Audited?

The Institute of Electrical and Electronics Engineers (IEEE) defines an audit for software "products and processes" as "an independent evaluation of conformance ... to applicable regulations, standards, guidelines, plans, specifications, and procedures."[123] An algorithmic process runs from specification of the problem through data collection, modeling, and validation to deployment and even post-deployment adjustments. For dynamic processes, like social media algorithms, this process is iterative and constantly renewing. Algorithmic auditing provisions using terms like "risk assessment" or "audit" are often vague about the object and timing of the inquiry, and whether they intend to look at the full life cycle of an AI system or only parts of it.

Some audits will focus on code. When Elon Musk announced that he would make Twitter's algorithm "open source" if he owned the platform, the promise was that its content ranking decisions would be subject to review.[124] Critics responded that code alone does not make algorithms legible and accountable.[125] The compute and training data at the technical core of algorithmic functions are important foci for any review. But so are the complex human and sociotechnical choices that shape the algorithmic process, including human selection of objectives and override of algorithmic recommendations. An open-source code does not necessarily enable others to replicate results, much less explain them.[126] Varied kinds and levels of information are appropriate depending on who wants to know what, and also on the necessary degree of protection for proprietary information.

The what of an audit is inextricably tied to the when. What points of the algorithmic process are in view? If the goal of the audit is principally reflexive—that is to help developers catch problems and better inculcate a compliance mindset—then the audit should be forward-looking and implemented at early stages before deployment. Such an "audit" actually then functions like an algorithmic impact assessment. "An example of reflexive regulation, impact assessment frameworks are meant to be early-stage interventions, to inform projects before they are built," writes Andrew Selbst.[127] Canada's algorithmic impact assessment tool, for example, requires the inquiry to "be completed at the beginning of the design phase of a project ...[and] a second time, prior to the production of the system, to validate that the results accurately reflect the system that was built."[128] AI Now's framework for impact assessments, focusing on public accountability for the use of automated systems by public agencies, similarly looks at pre-deployment.[129] So too, the AI Act's conformity assessments are to be done pre-deployment for high-risk systems per Articles 16 and 43.[130]

By contrast, an audit designed to check whether a firm's product actually delivers on promises or complies with the law will be backward-looking as, for example, in the DSA's required audits of risk assessment and mitigation measures. Researcher access to data will also support lookback audits of already-deployed systems. A recent European Parliament report proposes incorporating into the AI Act individual transparency rights for subjects of AI systems, which also supports post-hoc review.[131] Because many algorithmic systems are incessantly dynamic, the distinction between ex post and ex ante may be exaggerated. Every look back is a look forward and can inform the modification of algorithmic systems, creating accountability for and prevention of algorithmic harm. The cyclical process of AI development and assessment

123 IEEE, Standard for Software Reviews and Audits, IEEE Std 1028-2008, August 15, 2008.

124 Maxwell Adler, "Why Elon Musk Wants to 'Open Source' Twitter's Algorithms," Blomberg, April 28, 2022.

125 Cathy O'Neil, "Sorry Elon, 'Open Source' Algorithms Won't Improve Twitter," The Washington Post, May 2, 2022.

126 Deven R. Desai and Joshua A. Kroll, "Trust but verify: A guide to algorithms and the law," Harvard Journal of Law & Technology, April 27, 2017.

127 Andrew D. Selbst, "An Institutional View of Algorithmic Impact Assessments," Harvard Journal of Law & Technology, June 24, 2021.

128 Treasury Board of Canada Secretariat, Algorithmic Impact Assessment tool.

129 Dillon Reisman et al., Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability, AI Now Institute, April 2018.

130 European Commission, Proposal for a Regulation of The European Parliament and The Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, April 21, 2021.

131 Panel for the Future of Science and Technology (STOA), Auditing the quality of datasets used in algorithmic decision-making systems, European Parliament, July 2022.

shows up, for example, in how the US NIST conceptualizes the perpetuation of bias in AI, from pre-design in which "problem formulation may end up strengthening systemic historical and institutional biases" to design and development where "models based on constructs via indirect measurement with data reflecting existing biases" to deployment, wherein "heuristics from human interpretation and decision-making and biases from institutional practices."[132]

Whatever part of the process the audit examines, auditors will need records and audited entities will have to create relevant audit trails. Such trails, as Miles Brundage and co-authors write,

> could cover all steps of the AI development process, from the institutional work of problem and purpose definition leading up to the initial creation of a system, to the training and development of that system, all the way to retrospective accident analysis.[133]

Extending the audit trail beyond merely technical decisions would reflect how an algorithmic system fits into the larger sociotechnical context of an entity's decision-making. Focusing merely on software, as Mona Sloane has shown, fails to account for wider biases and underlying assumptions shaping the system.[134] Audits may require access not only to technical inputs and model features, but also to how teams are constituted, who makes decisions, how concerns are surfaced and treated, and other soft tissue elements surrounding the technical system. As Andrew Selbst and co-authors have cautioned, a narrowly technical audit will miss important framing decisions that dictate how an AI system functions and for what purpose.[135] Some biased outcomes may be further entrenched or perpetuated when the same datasets or models are deployed in algorithmic tools across multiple settings and by different actors. Audits may thus be an imperfect or less useful tool with potential blinds spots when it comes to how "algorithmic monoculture" leads to this outcome homogenization.[136]

In other words, auditors will need insight into the membership of the development team and the issues that are made salient. What sorts of outcomes does management want the AI system optimized for? What possibilities exist to override an AI system? What are the procedures for review and response to AI operations? Jennifer Cobe and co-authors recommend a "holistic understanding of automated decisionmaking as a broad sociotechnical process, involving both human and technical elements, beginning with the conception of the system and extending through to use consequences, and investigation."[137] Transparency around or audits of code alone will not be sufficient to reveal how algorithmic decisionmaking is happening. Furthermore, lab tests provide incomplete and possibly misleading reassurance. A particular algorithmic system may pass a lab test but not perform adequately in the "wild." Lab success or failures supply meaningful data points but should not stand in for audits of systems as they are practiced.[138]

Why: What Are the Audit's Objectives?

The functional purpose of an audit can vary widely. An audit may serve as an adjunct to law enforcement, such as a government agency's conduct of an audit as part of an investigation.[139] Alternatively, an audit may entail private internal

132 NIST, A proposal for Identifying and Managing Bias in Artificial Intelligence, Special Publication 1270, June 2021.

133 Miles Brundage et al., "Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims," arXiv, April 20, 2020.

134 Mona Sloane, Emanuel Moss, and Rumman Chowdhury, "A Silicon Valley love triangle: Hiring algorithms, pseudo-science, and the quest for auditability," Patterns 3, no. 2, February 11, 2022. p. 3.

135 Andrew D. Selbst et al., "Fairness and Abstraction in Sociotechnical Systems," FAT* '19, Proceedings of the Conference on Fairness, Accountability, and Transparency, January 2019, p. 59.

136 Comment by Deborah Raji in October 26, 2022 GMF workshop. See Github, "Picking on the same person: Does Algorithmic Monoculture lead to Outcome Homogenization," September 20, 2022. Outcome homogenization is "the extent to which particular individuals or groups experience the same outcomes across different deployments."

137 Cobbe, Lee, and Singh, "Reviewable Automated Decision-Making," p. 599.

138 Comments by Solon Barocas and Deborah Raji in October 26, 2022 GMF-RIIPL workshop. See Aaron Reike et al, "Essential Work: Analyzing the Hiring Technologies of Large Hourly Employees," Upturn, July 6, 2021; "Participatory Data Stewardship: A framework for involving people in the use of data," Ada Lovelace Institute, September 7, 2021.

139 Government of Canada, Summary of Session Four: Regulatory Powers, Report of Expert Advisory Group on Online Safety, May 6, 2022.

or external reviews of algorithmic functions to demonstrate compliance with an ethical or legal standard or to provide assurance that the algorithm functions as represented. Audit provisions should answer the question of why audit.

One of the most broadly accepted purposes of an audit is to signal compliance with, or at least consideration of, high-level ethical guidelines. There are many codes of ethics propounded for AI. Brent Mittelstadt surveyed the field in 2019 and found at least 84 AI ethics initiatives publishing frameworks.[140] Another fruitful source of objectives is the UN Guiding Principles Reporting Framework, which provides human rights-related goals for businesses, and is the metric that Meta has used to audit its own products.[141] Yet another potentially influential set of objectives emerges from the 2019 Ethics Guidelines for Trustworthy AI published by the European Commission's High-Level Expert Group on AI.[142] While research has shown that high-level ethical guidelines have not influenced the behavior of software engineers in the past,[143] it remains to be seen whether audit practices could help operationalize ethical principles for engineers of the future.

Whether framed as an ethical goal or a legal requirement, the functional objectives for algorithmic audits often fall into the following categories:

- **Fairness.** The audit checks whether the system is biased against individuals or groups vis-à-vis defined demographic characteristics.
- **Interpretability and explainability.** The audit checks whether the system makes decisions or recommendations that can be understood by users and developers, as is required in the GDPR.
- **Due process and redress.** The audit checks whether a system provides users with adequate opportunities to challenge decisions or suggestions.
- **Privacy.** The audit checks whether the data governance scheme is privacy-protecting and otherwise compliant with best practices.
- **Robustness and security.** The audit checks that a system is operating the way it is "supposed to" and is resilient to attack and adversarial action.

For social media platform governance in particular, audit advocates frequently point to bias, explainability, and robustness as objects of inquiry. Civil society wants assurance that service providers are moderating and recommending content in ways that do not discriminate, that are transparent, and that accord with their own terms of service.[144] Meta has now conducted a human rights audit itself,[145] but resisted submitting to external audits. Other inquiries relate to how platforms course-correct when new risks arise. The DSA and draft UK Online Safety bill include auditing provisions for mitigation. A related question concerns how algorithmic and human systems work together—that is, how are the systems structured to respond to concerns raised by staff or outside members of the public?

With respect to any given function, such as privacy, security, or transparency, auditing frameworks can differ in how they organize the inquiry. The Netherlands, for example, has set forth an auditing framework for government use of algorithms organized along the lines of management teams.[146] First, it looks at "governance and accountability." This inquiry focuses on the management of the algorithm throughout its life, including who has what responsibilities and where liability lies. Second, it looks at "model and data," examining questions about data quality, and the development, use, and maintenance of the model underlying the algorithm. This would include questions about bias, data

140 Brent Mittelstadt, "Principles Alone Cannot Guarantee Ethical AI," Nature Machine Intelligence 1, November 2019.

141 Shift, UN Guiding Principles Reporting Framework, accessed July 24, 2022.

142 European Commission, Ethics Guidelines for Trustworthy AI, accessed July 24, 2022.

143 Thilo Hagendorff, "The ethics of AI ethics: An evaluation of guidelines," Minds and Machines, February 1, 2020.

144 The Aspen Institute, Commission on Information Disorder Final Report, November 15, 2021; Rebecca Heilweil, "Facebook is taking a hard look at racial bias in its algorithms," Vox, July 22, 2020.

145 Miranda Sissons and Ian Levine, A Closer Look: Meta's First Annual Human Rights Report, Meta, July 14, 2022.

146 Netherlands Court of Audit, Understanding Algorithms, p. 24.

minimization, and output testing. Third, it looks at privacy, including compliance with GDPR. Fourth, it examines "information technology general controls." These concern management of access rights to data and models, security controls, and change management. Having adopted this audit framework, the Netherlands Court of Audit went on to find that only three of nine algorithms it audited complied with its standards.[147]

Whatever the audit objective and structure, mere assessment without accountability will not accomplish what audit proponents promise. As Mike Ananny and Kate Crawford have written, accountability "requires not just seeing inside any one component of an assemblage but understanding how it works as a system."[148] Sasha Costanza-Chock and co-authors recommend that the applicable accountability framework be explicitly defined.[149] An audit that seeks to measure compliance with human rights standards, for example, must identify the applicable equality or privacy norms and then how those norms have or have not been operationalized. There must also be a structure for imposing consequences for falling short.

Finally, addressing the question of "why audit" requires consideration of potential attendant costs.[150] Scholars have criticized audits for tacitly accepting the underlying assumptions of tools such as hiring algorithms, thereby seeming to validate pseudo-scientific theories that may have given rise to the tool.[151] In this way, audits may risk legitimizing tools or systems that perhaps should not exist at all. In addition, auditing processes may also require an entity to divert limited resources from innovation, which may impair the ability of new entrants and smaller firms, in particular, to compete. Auditing as a regulatory tool can also entail governance costs. The very project of auditing, to the extent that it involves government, may blur a public-private distinctions, bringing government into private processes. When audits become a preferred regulatory approach, whatever standard is audited to can become the ceiling for performance—businesses are encouraged to satisfy a measurable standard, which becomes ossified and perhaps below what entities might otherwise achieve by making different kinds of investments. Those subject to audit may be reluctant to discover or share information internally out of concern that it will hurt them in an audit, and this difficult-to-quantify chilling effect may also engender downstream costs. The benefits of audits may well justify these costs, but they should be considered.

How: Audit Standards

Imprecision or conflicts in audit standards and methodology within or across sectors may make audit results at best contestable and at worst misleading. "As audits have proliferated..., the meaning of the term has become ambiguous, making it hard to pin down what audits actually entail and what they aim to deliver," write Briana Vecchione and co-authors.[152] Some of this difficulty stems from the lack of agreed methods by which an audit is conducted. The question of how an audit is conducted may refer to "by what means" it is conducted, or it may refer to "by what standards" it is conducted.

UK regulators have addressed the means question, categorizing audit techniques as: technical audits that look "under the hood" at system components such as data and code; empirical audits that measure the effects of an algorithmic system by examining inputs and outputs; and governance audits that assess the procedures around data use and decision architectures.[153] The Ada Lovelace Institute

147 Netherlands Court of Audit, An Audit of 9 Algorithms used by the Dutch Government, May 18, 2022.

148 Ananny and Crawford, "Seeing without Knowing," p. 983.

149 Sasha Costanza-Chock, Inioluwa Deborah Raji, and Joy Buolamwini, "Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem," FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency, June 2022, p. 1580.

150 Inspired by observations by Niva Elkin-Koren, professor of law, Tel Aviv University, during GMF workshop, October 26, 2022.

151 Mona Sloane, Emanuel Moss, and Rumman Chowdhury, "A Silicon Valley Love Triangle: Hiring algorithms, pseudo-science, and the quest for auditability," Patterns 2, no. 2, February 2022, p. 25; Rhea, Alene, et al. "Resume Format, LinkedIn URLs and Other Unexpected Influences on AI Personality Prediction in Hiring: Results of an Audit," Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society, 2022.

152 Briana Vecchione, Solon Barocas, and Karen Levy, "Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies," EAAMO'21: Equity and Access in Algorithms, Mechanisms, and Optimization, no. 19, October 2021, p. 1.

153 Government of the United Kingdom, Auditing Algorithms.

Case Study: Washington, DC Stop Discrimination by Algorithms Act of 2021

A proposed Washington DC regulation, the Stop Discrimination by Algorithms Act of 2021, would require algorithmic auditing by businesses making or supporting decisions on important life opportunities.[1] The regulation specifies prohibited discriminatory practices to ensure that algorithmic processes comply with ordinarily applicable civil rights law. It charges businesses with self-auditing and reporting their findings. In this context, where the substantive standards (disparate impact) are clear, self-audit to those standards might be sufficient. The same approach in areas where the harms are less well understood or regulated will have different effects.

The law is concerned with algorithmic discrimination based on protected traits in the providing of access to or information about important life opportunities, including credit, education, employment, housing, public accommodation, and insurance. At the core of the law is a substantive prohibition (Section 4): "A covered entity shall not make an algorithmic eligibility determination or an algorithmic information availability determination on the basis of an individual's [protected trait]." This provision seeks to harmonize algorithmic practices with the protections of Washington DC's Human Rights Act of 1977. There is also a transparency provision (Section 6), which requires covered entities to provide notice of their use of personal information in algorithmic practices and notices and explanations of adverse decisions.

The audit provision (Section 7) builds up from the substantive and transparency requirements:

- Covered entities must do annual audits, consulting with qualified third parties, to analyze disparate-impact risks of algorithmic eligibility and information availability determinations.
- They must create and maintain audit trail records for five years for each eligibility determination including data inputs, algorithmic model, tests of model for discrimination, methodology for decision.
- They must also conduct annual impact assessments of existing algorithmic systems (backward looking) and new systems prior to implementation (forward-looking). These impact assessments are also referred to as "audits."
- The covered entities must implement a plan to reduce the risks of disparate impact identified in the audits.
- They then must submit an annual report to the Washington DC attorney general containing information about their algorithmic systems (what types of decisions they make, methodologies and optimization criteria used, upstream training data and modeling methodology, downstream metrics used to gauge algorithmic performance), information about their impact assessments and responses, and information about complaints and responses.

The who of the audit is the business itself. First-party audits are generally not going to be as trustworthy. In this case, some of the risks are mitigated by reporting out the results and methodology to the attorney general. This approach puts the onus on the government to be able to assess audit methodology.

The what of the audit includes upstream inputs to the algorithmic model and its outputs. It does not seem to include the humans in the loop or other non-technical features of the algorithmic decision-making.

The why is very clear in part because the civil rights standards of wrongful discrimination are well-established, and the practice is prohibited. The how is entirely unspecified. Covered entities can choose how they conduct audits, with the only check being that they are supposed to report their methodology to the Attorney General. These reports are not made public, at least in the first instance.

1 Council of the District of Columbia, Stop Discrimination by Algorithms Act of 2021, proposed December 8, 2021.

Case Study: Netherlands Audit of Public Algorithms

The example of the Netherlands' audit of public algorithms answers the what, why, and who questions about algorithmic audits fairly clearly. This is easier to do when the government itself is conducting the audits of systems that it controls. Even here, however, the how of the audit practice is not clear and so it is difficult to compare the findings to similar kinds of audits of other systems and in other jurisdictions.

In March 2022, the Dutch government released the results of an audit examining nine algorithms used in government agencies.[1] The audit found that three algorithms met the audit requirements, while six failed the audit. The topic is a hot-button one in the Netherlands following the 2019 *toeslagenaffaire*, or child benefits scandal, in which a government algorithm used to detect benefits fraud erroneously penalized thousands of families and placed over 1,000 children in foster care based on "risk factors" like dual nationality or low income.[2]

The audit was based on a framework laid out in the 2021 report *Understanding Algorithms* from the Netherlands Court of Audit.[3] The auditing framework is publicly available for download.[4] The framework assesses algorithms across five metrics: governance and accountability; model and data; privacy; IT general controls; and ethics, which encompasses respect for human autonomy, the prevention of damage, fairness, explicability, and transparency.

The audit was carried out according to the following questions:

1. Does the central government make responsible use of the algorithms that we selected?
   a. Have sufficiently effective controls been put in place to mitigate the risks?
   b. Do the algorithms that we selected meet the criteria set out in our audit framework for algorithms?
2. How do the selected algorithms operate in practice? How does each algorithm fit in with the policy process as a whole?
   a. How does the government arrive at a decision on the use of the algorithm?
   b. What do officials do with the algorithm's output? On which basis are decisions taken?
   c. What impact does this have on private citizens?[5]

The nine algorithms were selected according to the following criteria: impact on private citizens or businesses; risk-centered, or those with the highest risk of misuse; different domains or sectors; algorithms currently in operation; and different types, from technically simple algorithms such as decision trees to technically more complex algorithms like image recognition systems.[6]

Each agency audit was conducted by at least two auditors according to the audit framework and using documentation from the agencies, interviews, and observations. Audited agencies were asked to confirm outcomes of an assessment and provide complementary documentation and details before a reassessment. The overall assessment was made by the entire audit team.

The Dutch example is a useful illustration of an auditing framework in action, with a broad mandate to examine decision-making systems in everyday use. Its results are a clear example of the various ways in which risk can arise

1 Netherlands Court of Audit, An Audit of 9 Algorithms used by the Dutch Government.
2 Melissa Heikkila, "Dutch scandal serves as a warning for Europe over risks of using algorithms," Politico, March 29, 2022.
3 Netherlands Court of Audit, Understanding Algorithms.
4 Netherlands Court of Audit, Audit Framework for Algorithms, January 26, 2021.
5 Netherlands Court of Audit, An Audit of 9 Algorithms used by the Dutch Government, p. 42.
6 Ibid., p. 43.

in the use of an algorithm, from insecure IT practices, to outsourcing of government algorithms to outside actors, to data management policies. This framework could be used as a model for defining higher-level standards for auditing. Yet it has drawbacks as a directly applicable model for algorithmic audits generally. For example, private companies might provide less access to data and proprietary information than in this government-on-government audit. Private auditing firms would also need to meet standards or certification criteria laid out by a governing body or national regulator to ensure audit quality and necessary changes if an algorithm or firm fails.

has developed a taxonomy of social media audit methods, focusing on scraping, accessing data through application programming interfaces, and analyzing code.[154] By whatever means an audit is conducted, its conclusions will depend on its purpose (discussed above) and its standards.

For standards, the question is how to build common or at least clear metrics for achieving audit goals. The Mozilla Foundation observes that algorithmic audits are "surprisingly ad hoc, developed in isolation of other efforts and reliant on either custom tooling or mainstream resources that fall short of facilitating the actual audit goals of accountability."[155] Shea Brown and co-authors found that "current proposals for ethical assessment of algorithms are either too high level to be put into practice without further guidance, or they focus on very specific and technical notions of fairness or transparency that do not consider multiple stakeholders or the broader social context."[156] The UK's Centre for Data Ethics and Innovation has announced that it "will support the Department for Digital, Culture, Media and Sport (DCMS) Digital Standards team and the Office for AI (OAI) as they establish an AI Standards Hub, focused on global digital technical standards."[157] For the DSA, auditors like Deloitte are proposing to apply their own methodologies:

> The specific parameters and audit methodology required to produce the required [DSA] independent audit opinion have not been laid out in the Act and so firms and their chosen auditors will need to consider the format, approach and detailed methodology required to meet these requirements ahead of the audit execution.[158]

A common set of standards remains contested and elusive as the goals and basic definitions of both the auditors and the audited conflict.

The results of audits should allow interested parties to understand and verify claims that entities make about their systems. With respect to financial audits, US federal law authorizes the Securities and Exchange Commission (SEC) to set financial accounting standards for public companies and lets it recognize the standards set by an independent organization. The SEC has recognized standards adopted by the Financial Accounting Standards Board—a nonprofit consisting of a seven-person board—as authoritative.[159] In the tech context, a similar sort of co-regulation shapes Australia's Online Safety Act of 2021, the UK Online Safety Act, and the EU DSA, all of which make use of industry codes of conduct.[160] Codes of conduct, while of course not themselves audit standards, can be precursors to them. Audits can use codes to supply the "why" and "how" of an audit.

154 Ada Lovelace Institute, Technical methods for regulatory inspection of algorithmic systems, December 9, 2021.

155 Deb Raji, It's Time to Develop the Tools We Need to Hold Algorithms Accountable, Mozilla, February 2, 2022

156 Brown, Davidovic, and Hasan. "The Algorithm Audit," p. 1.

157 Government of the United Kingdom, The Roadmap to an Effective AI Assurance Ecosystem - Extended Version," accessed July 19, 2022.

158 Mark Cankett and Lenka Fackovcova, EU Digital Services Act: Are you ready for audit?, Deloitte, May 18, 2022.

159 US Securities and Exchange Commission, Policy Statement: Reaffirming the Status of the FASB as a Designated Private-Sector Standard Setter, modified April 25, 2003; Financial Accounting Standards Board (FASB), About the FASB, accessed July 23, 2022.

160 DSA Article 34(1)(d) explicitly mentions a potential "voluntary standard" for audits. For a list of possible "standards," see Julian Jaursch, Overview of DSA delegated acts, reports and codes of conduct, Stiftung Neue Verantwortung, August 15, 2022.

These codes might look like those being developed by the Partnership on AI, for example, which is creating codes of conduct for industry with respect to distinct problems like synthetic media and biometrics.[161] Still other standards will emerge from legacy standard-setting bodies, such as the IEEE, which has an initiative on Ethically Aligned Design.[162] In a 2019 report, this IEEE initiative said that "companies should make their systems auditable and should explore novel methods for external and internal auditing."[163] It included proposals for how to make information available to support audits by different stakeholders and for different purposes.

Miles Brundage and co-authors have proposed a number of specific recommendations for work by standards-setting bodies in conjunction with academia and industry to develop audit techniques.[164] Alternatively, government entities themselves might set standards. For example, the EU Expert Group on AI, which cited auditability as a key element of trustworthy AI systems in its 2019 ethics guidelines, is producing specific guidance for algorithmic audits in the financial, health, and communications sectors.[165]

Conclusion

Audits of automated decision systems, variously also called algorithmic or AI systems, are currently required by the EU's Digital Services Act, arguably by the EU's GDPR, and either required or considered in a host of US laws. Audits are proposed as a way to curb discrimination and disinformation, and to hold those who deploy algorithmic decision-making accountable for their harms. Many other uses of related terms, such as impact assessment, would also impose obligations on covered entities to benchmark the development and implementation of algorithmic systems against some acceptable standard.

For any of these interventions to work in the way that their proponents imagine, our review of the relevant provisions and proposals suggest that the term audit and associated terms require much more precision.

For any of these interventions to work in the way that their proponents imagine, our review of the relevant provisions and proposals suggest that the term audit and associated terms require much more precision.

Who. Key information about the person or organization expected to conduct the audit must be clear, including their qualifications and conditions of independence (if any), and their access to data and audit trails. If the audit is an internal one conducted by the covered entity itself, it should be clear how such an audit fits into a larger accountability scheme, and with guardrails in place to prevent algorithm-washing.

What. The subject of the audit should be explicit. The mere statement that a system should be audited leaves open the possibility of many different kinds of examinations, for example of models, of human decision-making around outputs, of data access and sharing. Even just taking the first example of a technical audit, the inquiry might focus on model development only or include system outputs, and also cover different periods. The range of audit scope expands further when one recognizes that the technical components of an algorithmic system are embedded in sociopolitical structures that affect how the technology works in context. Audit provisions should be clear about their scope.

Why. Audit objectives should also be specified. The ethical or legal norms with which an audit can engage are varied and sometimes conflicting. Whether the audit seeks to confirm compliance with a narrow legal standard or enquires about a broader range of ethical commitments, the goals should be transparent and well-defined. This is important not only intrinsically for any audit, but also for facilitating comparisons between audit findings. Specifying the purpose of the

161 Claire Leibowicz, PAI Developing Ethical Guidelines for Synthetic Media, Partnership on AI, March 10, 2022.

162 IEEE, Ethically Aligned Design, IEEE Ethics in Acton in Autonomous and Intelligent System, accessed July 23, 2022.

163 IEEE, Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, 1st Edition, 2019.

164 Brundage et al., "Toward Trustworthy AI Development," p. 3

165 High-Level Expert Group on Artificial Intelligence (AI HLEG), Ethics Guidelines for Trustworthy AI, European Commission, April 2019; AI HLEG, Sectoral considerations on policy and investment recommendations for trustworthy AI, European Commission, July 2020.

audit should also take account of the potential costs for the audited entity, the regulator (if any), and the public.

How. The standards the audit uses to assess norms like fairness, privacy, and accuracy should be as consensus-driven as possible. In the absence of consensus, which will be frequent, the standards being applied should be at minimum well-articulated. A situation in which auditors propose their own standards is not ideal. Common (or at least evident) standards will foster civil society's development of certifications and seals for algorithmic systems, while nebulous and conflicting standards will make it easier to "audit-wash" systems, giving the false impression of rigorous vetting.

As algorithmic decision systems increasingly play a central role in critical social functions—hiring, housing, education, and communication—the calls for algorithmic auditing and the rise of an accompanying industry and legal codification are welcome developments. But as we have shown, basic components and commitments of this still nascent field require working through before audits can reliably address algorithmic harms.

About the Author(s)

Ellen P. Goodman is senior advisor for algorithmic Justice at the National Telecommunications and Information Administration, distinguished professor at Rutgers Law School, and a former non-resident senior fellow at the German Marshall Fund of the United States. The views in this paper reflect her personal views and not those of the NTIA or any agency within the US government. Julia Tréhu is program manager and fellow with the Digital Innovation and Democracy Initiative at the German Marshall Fund of the United States.

Acknowledgments

The authors thank the participants in GMF Digital's October 26, 2022 Workshop on AI Audits as well as Julian Jaursch from the Stiftung Neue Verantwortung and Karen Kornbluh of the German Marshall Fund for comments on earlier drafts of the paper.

About GMF Digital

The German Marshall Fund's Digital Innovation and Democracy Initiative (GMF Digital) works to support democracy in the digital age. GMF Digital leverages a transatlantic network of senior fellows to develop and advance strategic reforms that foster innovation, create opportunity, and advance an equitable society.

Cover photo credit: metamorworks | Shutterstock

About GMF

The German Marshall Fund of the United States (GMF) is a non-partisan policy organization committed to the idea that the United States and Europe are stronger together. GMF champions the principles of democracy, human rights, and international cooperation, which have served as the bedrock of peace and prosperity since the end of World War II, but are under increasing strain. GMF works on issues critical to transatlantic interests in the 21st century, including the future of democracy, security and defense, geopolitics and the rise of China, and technology and innovation. By drawing on and fostering a community of people with diverse life experiences and political perspectives, GMF pursues its mission by driving the policy debate through cutting-edge analysis and convening, fortifying civil society, and cultivating the next generation of leaders on both sides of the Atlantic. Founded in 1972 through a gift from Germany as a tribute to the Marshall Plan, GMF is headquartered in Washington, DC, with offices in Berlin, Brussels, Ankara, Belgrade, Bucharest, Paris, and Warsaw.