Santa Clara High Technology Law Journal

Volume 39 | Issue 3 Article 1

5-26-2023

ALGORITHMIC AUDITING: CHASING AI ACCOUNTABILITY

Goodman, Ellen P.

Trehu, Julia

ALGORITHMIC AUDITING: CHASING AI ACCOUNTABILITY

Ellen P. Goodman[*]
Julia Trehu[•]

[*] Distinguished Professor, Rutgers Law School and Senior Advisor for Algorithmic Justice, National Telecommunications and Information Administration. The views herein are the author's own and do not reflect any official position. Both authors thank participants of 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.

[•] Program Manager and Fellow, Digital Innovation and Democracy Initiative, German Marshall Fund.

CONTENTS

"Algorithmic auditing refers to a range of approaches to review algorithmic processing systems. It can take different forms, from checking governance documentation, to testing an algorithm's outputs, to inspecting its inner workings."[1]

"Algorithm auditing is the research and practice of assessing, mitigating, and assuring an algorithm's safety, legality, and ethics."[2]

Calls for audits to expose and mitigate harms related to algorithmic decision systems are proliferating,[3] and audit provisions are coming into force—notably in the E.U. Digital Services Act.[4] In response to these growing concerns, 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.[5] 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.[6] Even when not expressly prescribed, audits may be predicates for enforcing data-related consumer protection law, or what U.S. Federal Trade Commissioner Rebecca Slaughter calls

---

[1] *Auditing Algorithms: The Existing Landscape, Role of Regulators and Future Outlook*, DIG. REGUL. COOP. F. (last updated Sept. 23, 2022).

[2] Adriano Koshiyama et al., *Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms*, SSRN ELEC. J. 1, 2 (Jan. 2021).

[3] Shea Brown et al., *The Algorithm Audit: Scoring the Algorithms That Score Us*, 8 BIG DATA & SOC'Y 1, 1 (2021) ("In response to these growing concerns, nearly every research organization that deals with the ethics of AI has called for ethical auditing of algorithms.").

[4] Press Release, *Digital Services Act: Commission Welcomes Political Agreement on Rules Ensuring a Safe and Accountable Online Environment*, EU COMM'N (Apr. 23, 2022).

[5] *See, e.g.*, *Achieving trustworthy AI: A Model for Trustworthy Artificial Intelligence*, KPMG AUSTL. (Nov. 24, 2020); *Deloitte Introduces Trustworthy AI Framework to Guide Organizations in Ethical Application of Technology in the Age of With*, DELOITTE (Aug. 2020).

[6] *See, e.g.*, *The Use of Artificial Intelligence and Machine Learning by Market Intermediaries and Asset Managers*, INT'L ORG. SEC. COMM'N (June 2020); *SR 11-7: Guidance on Model Risk Management*, BD. GOVERNORS FED. RSRV. SYS. (Apr. 04, 2011) (requiring audits of machine learning models); Laurent Dupont et al., *Governance of Artificial Intelligence in Finance*, ACPR BANQUE DE FR. (June 2020) (discussing AI auditing requirements in financial regulation).

"algorithmic justice."[7] The desire for audits reflect

> 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.[8]

Much as financial audits transformed the way businesses operated in the twentieth century, algorithmic audits can transform the way technology works in the twenty-first. Stanford University's 2022 AI Audit Challenge lists the benefits of AI auditing, namely verification, performance, and governance. AI Auditing

> 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.[9]

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

---

[7] Rebecca Kelly Slaughter et al., *Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission*, 23 YALE INFO. SOC'Y PROJECT & YALE J.L. & TECH. 1, 56 (2021) (explaining algorithmic justice entails civil rights protections to "limit the dangers of algorithmic bias and require companies to be proactive in avoiding discriminatory outcomes").

[8] Briana Vecchione et al., *Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies*, 2021 EQUITY & ACCESS ALGORITHMS, MECHANISMS & OPTIMIZATION 1–9 (Oct. 2021).

[9] Marietje Schaake & Jack Clark, *Stanford Launches AI Audit Challenge*, STAN. HAI: L., REGUL. & POL'Y (July 11, 2022).

frameworks. To this end, one must confront the reality that audit discourse tends to be inexact and confusing.[10] 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 measures open up "black box" algorithms to public scrutiny and then audits are conducted once the lid is off.[11] 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.[12]

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*. We recognize that the definition of "algorithm" is broad and distinct from the definition of AI, since not all algorithms use AI.[13] But audit provisions have as their central concern an AI process—defined by the U.S. National Artificial Intelligence Initiative Act of 2020 as, "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments."[14] Therefore, we use the terms

---

[10] *See* Jacqui Ayling & Adriane Chapman, *Putting AI Ethics to Work: Are the Tools Fit for Purpose?*, 2021 PROC. EQUITY & ACCESS ALGORITHMS, MECHANISMS, & OPTIMIZATION, AI & ETHICS 405, 421 (Sept. 12, 2021) (illustrating "the confusion in language and approach to what are understood as the key features of impact assessment and audit"); *see also* Ghazi Ahamat et al., *Types of Assurance in AI and the Role of Standards*, CTR. DATA ETHICS & INNOVATION BLOG (Apr. 17, 2021) ("The current discourse sometimes mistakenly calls on risk assurance tools like impact assessments to achieve the goals of [c]ompliance[.] Meanwhile, sometimes compliance mechanisms like audits are discussed as if they can achieve loftier goals—an exercise which may be better suited to Risk Assurance tools like impact assessments.").

[11] Tom Cassauwers, *Opening the Black Box of Artificial Intelligence*, HORIZON—THE E.U. RSCH. & INNOVATION MAG. (Dec. 01, 2020).

[12] James Guszcza et al., *Why We Need to Audit Algorithms*, HARV. BUS. REV: ANALYTICS & DATA SCI. (Nov. 28, 2018) ("Companies have long been required to issue audited financial statements for the benefit of financial markets and other stakeholders . . . . [i]ndependent auditors are hired to provide reasonable assurance that the reports coming from the 'black box' are free of material misstatement . . . . [e]conomically independent bodies could be formed to deliberate and issue standards of design, reporting and conduct.").

[13] Kristian Lum & Rumman Chowdhury, *What is An 'Algorithm'? It Depends on Whom You Ask*, MIT TECH. REV. (Feb. 26, 2021).

[14] William M. (Mac) Thornberry, Nat'l Def. Auth. Act for Fiscal Year 2021, Pub. L. No. 116-283, § 5002, 134 Stat. 3388; *see also* Nat'l A.I.

"AI" and "algorithmic" audit interchangeably without insisting on any particular definition of these terms.

In posing these questions, we do not suggest that audits will look the same within a sector or across sectors. Audits of high-risk systems, such as biometric sorting in law enforcement,[15] will be different from audits of lower-risk systems, such as office utilization detection in property management.[16] The European Union's proposed AI Act distinguishes among risk categories for audit and other purposes,[17] and we suspect the future of audit regulation will be strongly inflected with this approach. 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.[18] What an audit "sees" depends on what aspects of this complex

Initiative, 15 U.S.C.A. § 9401(3) (defining AI as software that is developed with techniques and approaches which can, "for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with"); Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap,* 51 U.C. DAVIS L. REV. 399, 404 (2017) (defining AI as a "set of techniques aimed at approximating some aspect of human or animal cognition using machines").

[15] Kashmir Hill, *Another Arrest, and Jail Time, Due to a Bad Facial Recognition Match*, N.Y. TIMES (Dec. 29, 2020).

[16] Patrick Sisson, *How Data Is Changing the Way Offices Are Run*, N.Y. TIMES (Apr. 27, 2021).

[17] *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*, COM (2021) 206 final (Apr. 21, 2021).

[18] Joshua A. Kroll, *Responsible AI is a Management Problem, not a Purchase*, REGUL. REV. (July 04, 2022).

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.[19] 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-specific regulations, or particularized measures of fairness, non-discrimination, and data protection.[20] Another audit objective might be to assure stakeholders that the system functions as represented, and 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

[19] Jennifer Cobbe et al., *Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems*, PROC. 2021 ACM CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 598, 599 (Mar. 2021) ("We set out a holistic understanding of [algorithmic 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."); Jacob Mökander & Maria Axente, *Ethics-based auditing of automated decision-making systems: intervention points and policy implications*, AI & SOC'Y 1, 1 (2021) (segmenting the AI lifecycle into organization, concept, development, evaluation and operation, and urging that auditing span all phases).

[20] Lorna McGregor et al., *International Human Rights Law as a Framework for Algorithmic Accountability*, 68 INT'L & COMPAR. L.Q. 309, 311 (2019) (discussing algorithmic accountability as a means of ensuring transparency, explainability, understandability, and protection of human rights); Margot E Kaminski & Gianclaudio Malgieri, *Algorithmic impact assessments under the GDPR: Producing Multi-Layered Explanations*, 11 INT'L DATA PRIV. L. 125, 125 (2021) (discussing how an algorithmic impact assessment is meant to provide systemic governance and safeguard individual rights).

of either the compliance or assurance audit is to create more reflexive internal processes around the development and deployment of AI systems.[21] The audit's objectives will have a 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, both for the audited entity or regulatory agencies.

- *How* is the audit being conducted? The methodology and standards by which the audit is conducted will affect its legitimacy.[22] 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.

I. ALGORITHMIC AUDITS—ACCOUNTABILITY OR FALSE ASSURANCE

Algorithmic audits can potentially address two related

---

[21] *See, e.g.*, Bogdana Rakova et al., *Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices*, 2021 PROC. ACM ON HUM.-COMPUT. INTERACTION 5 (2021) (AI system practitioners want AI reviews to be conducted prior to release of new features and audit teams to be integrated with machine learning operations.); Mökander & Axente, *supra* note 19 ("The main function of [ethics based auditing] should . . . be to inform, formalize, assess, and interlink existing governance structures.").

[22] *See, e.g.*, Adriano Koshiyama et al., *Familiar methods can help to ensure trustworthy AI as the algorithm auditing industry grows*, OECD AI POL'Y OBSERVATORY (Aug. 10, 2021).

problems: the opacity of machine learning algorithms and the illegal or unethical performance of algorithmic systems.[23] At the same time, audits can function as window-dressing, concealing fundamental social and technical deficiencies through false assurance.

A. *Accountability*

Concern has been growing over what Frank Pasquale called in his 2016 pathbreaking book "The Blackbox Society."[24] 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.[25] 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 Dutch government provides the following examples of different algorithms:

---

[23] *See* Inioluwa Deborah Raji et al., *Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing*, PROC. 2020 CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 33, 38 (Jan. 2020) (calling for an audit practice to "increase ethical foresight"); James Guszcza et al., *supra* note 12 (advocating for algorithmic audits as a means to ensure that AI technologies broadly reflect societal values); Shlomit Yanisky-Ravid & Sean K. Hallisey, *Equality and Privacy by Design: A New Model of Artificial Intelligence Data Transparency Via Auditing, Certification, and Safe Harbor Regimes*, 46 FORDHAM URB. L.J. 428, 429 (2019) (proposing "an auditing regime"); Bryan Casey et al., *Rethinking Explainable Machines: The GDPR's Right to Explanation Debate and the Rise of Algorithmic Audits in Enterprise*, 34 BERKELEY TECH. L.J. 143, 152 (2019) (discussing "data auditing methodologies" for promoting compliance with privacy laws); Christian Sandvig et al., *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*, 64TH ANN. MEETING INT'L COMMC'N ASS'N 1, 8 (May 22, 2014) (discussing audits as a way of reducing bias); *see also* Cary Coglianese & Erik Lampmann, *Contracting for Algorithmic Accountability*, 6 ADMIN. L. REV. ACCORD 194 (2021) (advocating for independent audits of government contractor AI systems).

[24] *See generally* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2016).

[25] *See generally* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L.J. 1085 (2018).

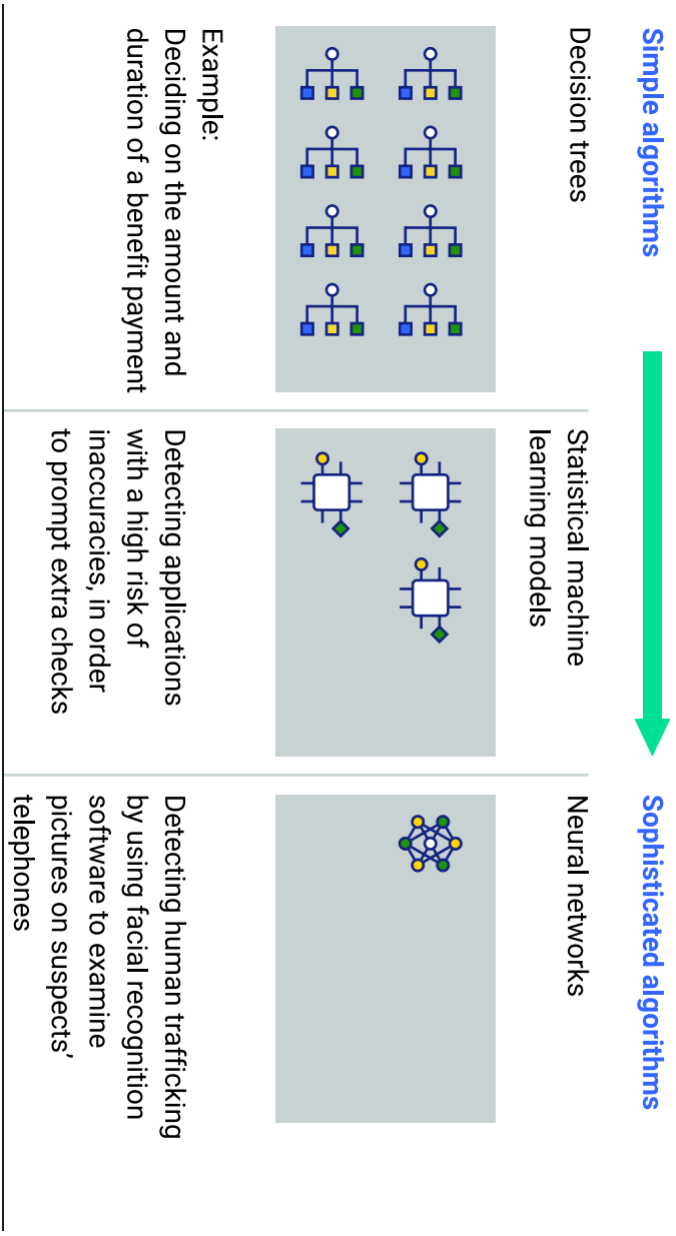

Figure 1 [26]

Opacity concerns are especially acute as the algorithmic process becomes more dependent on machine learning models.

[26] *Understanding Algorithms*, NETH. CT. AUDIT 18 (Jan. 26, 2021), www.english.rekenkamer.nl/publications/reports/2021/01/26/understanding-algorithms.

Particularly when they are used to inform critical determinations such as who gets hired[27] or policed,[28] the opacity of these processes can compromise public trust and accountability[29] and make it more difficult to challenge or improve decision making.[30]

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.[31] Social media algorithms can accelerate and broaden the spread of harmful information.[32] Algorithms involved in workplace productivity[33] and educational performance[34] 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.

It is notoriously difficult to regulate technology for many

---

[27] *See generally* Ifeoma Ajunwa, *The Auditing Imperative for Automated Hiring*, 34 HARV. J.L. & TECH. 621 (2021).

[28] *See generally* Elizabeth E. Joh, *Feeding the Machine: Policing, Crime Data, & Algorithms*, 26 WM. & MARY BILL RTS. J. 287 (2017).

[29] *See generally* Teresa M. Harrison & Luis Felipe Luna-Reyes, *Cultivating Trustworthy Artificial Intelligence in Digital Government*, 40 SOC. SCI. COMPUT. REV. 494 (2022); Cynthia Dwork & Martha Minow, *Distrust of Artificial Intelligence: Sources & Responses from Computer Science & Law*, 151 DAEDALUS 309 (2022); Baobao Zhang & Allan Dafoe, *U.S. Public Opinion on the Governance of Artificial Intelligence*, 2020 PROC. AAAI/ACM CONF. ON AI, ETHICS & SOC'Y (2020).

[30] *See* Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1964, 1965 (2021) (arguing for the importance of contestable AI).

[31] SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION 24 (2018); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 674 (2016); Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 875 (2017).

[32] *See Content-Sharing Algorithms, Processes, and Positive Interventions Working Group Part 1: Content-Sharing Algorithms & Processes*, GLOB. INTERNET F. TO COUNTER TERRORISM (July 2021), https://gifct.org/wp-content/uploads/2021/07/GIFCT-CAPI1-2021.pdf; Florian Saurwein & Charlotte Spencer-Smith, *Automated Trouble: The Role of Algorithmic Selection in Harms on Social Media Platforms*, 9 MEDIA & COMMC'N 222, 223 (2021); Wen-Ying Sylvia Chou & Anna Gaysynsky, *A Prologue to the Special Issue: Health Misinformation on Social Media*, 110 AM. J. PUB. HEALTH S270 (2020).

[33] Jodi Kantor & Arya Sundaram, *The Rise of the Worker Productivity Score*, N.Y. TIMES (Aug. 14, 2022).

[34] Amany Elbanna & Jostein Engesmo, *A-Level Results: Why Algorithms Get Things So Wrong – and What We Can Do to Fix Them*, CONVERSATION (Aug. 19, 2020).

reasons, including lack of institutional capacities[35] and the likelihood that technological change outpaces regulatory process.[36] Insisting on more transparency around the design and performance of algorithms is one response to the opacity problem.[37] Methods to force greater transparency include conducting algorithmic impact statements,[38] requiring researcher access to data,[39] and making aspects of government algorithmic systems transparent through records requests.[40] It must be recognized, however, that transparency alone is of limited utility for complex algorithmic systems.[41] 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.[42] The model will then use these learnings to make predictions about new scenarios.[43] How the model converts learnings into predictions is not

---

[35] *See, e.g.*, Rebecca Crootof & BJ Ard, *Structuring Techlaw*, 34 HARV. J.L. & TECH. 347, 376 (2021) (including institutional uncertainties among other challenges for the regulation of technology).

[36] *See generally* Gary E. Marchant, *The Growing Gap Between Emerging Technologies and the Law, in* 7 THE GROWING GAP BETWEEN EMERGING TECHNOLOGIES AND LEGAL-ETHICAL OVERSIGHT: THE PACING PROBLEM (Gary Marchant, Braden Allenby, & Joseph Herkert eds., 2011).

[37] Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J.L. & TECH. 103, 121, 129, 133 (2018).

[38] *See generally* Andrew D. Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J.L. & TECH. 117 (2021).

[39] *See generally* Nathaniel Persily, *A Proposal for Researcher Access to Platform Data: The Platform Transparency and Accountability Act*, 1 J. ONLINE TR. & SAFETY 1 (Oct. 10, 2021).

[40] *See* Hannah Bloch-Wehba, *Access to Algorithms*, 88 FORDHAM L. REV. 1265, 1296 (2020) (the "public interest in understanding how proprietary algorithmic governance works is precisely what is protected by laws requiring public access to government records and proceedings").

[41] *See* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 657–60 (2017) (detailing ways in which algorithms cannot be made legible).

[42] *See* Gabriel Nicholas, *Explaining Algorithmic Decisions*, 4 GEO. L. TECH. REV. 711, 714 (2020) ("Machine learning uses one algorithm, a *learner*, to output another algorithm that makes predictions, a *model*. The learner reads in data as a set of numerical features, infers rules about those features that predict the desired value, and outputs a model that embodies those rules.").

[43] *See* Brauneis & Goodman, *supra* note 37, at 113–14 (explaining an algorithmic process will typically involve "(1) the construction of a model to achieve some goal, based on analysis of collected historical data; (2) the coding of an algorithm that implements this model; (3) collection of data about subjects to provide inputs for the algorithm; (4) application of the prescribed

easy to render transparent.[44] The mere production of computer code or model features will be insufficient to make transparency meaningful.[45] The goal of making an algorithm legible to humans is now often expressed in terms of explainability[46] or interpretability,[47] rather than transparency. To this end, computer scientists are working in partnership with others to create "explainable AI" or "xAI."[48] Yet, so far at least, aspirational explainability cannot be relied upon either for effective communication about how algorithmic systems work or for holding them accountable.[49]

If well-designed and implemented, audits can abet transparency and explainability.[50] They can make visible aspects of system construction and operation that would otherwise be hidden.

---

algorithmic operations on the input data; and (5) outputs in the form of predictions or recommendations based on the chain of data analysis").

[44] *See, e.g.*, Katherine J. Strandburg, *Rulemaking and Inscrutable Automated Decision Tools*, 119 COLUM. L. REV. 1851, 1862 (2019) (explaining some models "map input features to outcome variables [in ways that] cannot be represented" for human comprehension); David Freeman Engstrom & Daniel E. Ho, *Algorithmic Accountability in the Administrative State,* 37 YALE J. REGUL. 800, 821 (2020); Brent Mittelstadt et al., *Explaining Explanations in AI*, 2019 PROC. CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY 279 (2019); Mike Ananny & Kate Crawford, *Seeing without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC'Y 973, 981 (2018).

[45] *See* Maayan Perel & Niva Elkin-Koren, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement*, 69 FLA. L. REV. 181, 184 (2017); Cansu Safak & Imogen Parker, *Meaningful Transparency and (In)visible Algorithms*, ADA LOVELACE INST. (Oct. 15, 2020); Matthew Gooding, *Elon Musk's plan for an open-source algorithm won't solve Twitter's problems*, TECH MONITOR (Apr. 26, 2020).

[46] *See* Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829, 1834 (2019) (defining explainable AI); Engstrom & Ho, *supra* note 44, at 824.

[47] Cynthia Rudin, *Stop Explaining Black Box Models for High Stakes Decisions and Use Interpretable Models Instead*, 1 NATURE MACH. INTELL. 206 (2019).

[48] *See generally* P. Jonathan Philips, et al., *Four Principles of Explainable Artificial Intelligence,* NAT'L INST. SCI. & TECH. 8312 (2021); David Gunning et al., *DARPA's Explainable AI (XAI) Program: A Retrospective*, WILEY ONLINE LIBR.: APPLIED AI LETTERS (Dec. 04, 2021).

[49] For a critique of xAI, *see, e.g.*, Nicholas, *supra* note 42, at 729 (xAI "cannot elucidate the real, internal reasons" for why systems produce certain results).

[50] *See* Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189, 190 (2017) ("Auditing is another method for forcing transparency.").

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.[51] This investigation can address the black box problem by providing assurance that the algorithm is working the way it is supposed to (e.g., accurately) and/or that it is compliant with applicable standards (e.g., 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.

B. *False Assurance*

Experience with audits in other contexts raises the specter of false assurance. A firm that has audited itself or submitted to inadequate audit 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.[52] Audit washing is a cousin of "greenwashing" and "ethics washing"—the acquisition of sustainability or ethical credibility through cosmetic or trivial steps.[53]

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."[54] Another common pitfall in the technology space is for a firm to profess adherence to human rights

---

[51] Because audits require the production of audit trails as a model is developed and deployed, investigation may entail disclosure/explanation. *See* Danielle Keats Citron, *Technological Due Process*, 85 WASH. U.L. REV. 1249, 1305 (2008) (discussing audit trails for public algorithms); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 121–24 (2014) (discussing audit trails for private-sector algorithms).

[52] Julian Jaursch, *Why The EU Needs To Get Audits For Tech Companies Right*, TECHDIRT (Aug. 19, 2021, 7:55 PM).

[53] *See, e.g.*, Elettra Bietti, *From Ethics Washing to Ethics Bashing: A Moral Philosophy View on Tech Ethics*, 2 J. SOC. COMPUTING 266 (2021).

[54] Andrew D. Selbst, *supra* note 38, at 144–45 (citing Mark Latonero & Aaina Agrawal, *Human Rights Impact Assessments For AI: Learning From Facebook's Failure In Myanmar*, CARR CTR. HUM. RIGHTS POL'Y (Mar. 19, 2021)).

standards without actually designing its systems to deliver on them.[55]

Even when outside checks are ostensibly in place, systems of assurance may simply mask wrongdoing. The U.S. 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.[56] An assessment is a less rigorous form of review than an 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."[57]

1. 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 Congress and civil rights groups, in 2018 Facebook (now Meta) commissioned a civil rights audit led by Laura Murphy, a former ACLU official, and Megan Cacace, a partner at Relman Colfax. They released a series of reports culminating in an eighty-nine-page audit report in July 2020.[58]

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 then-President Donald Trump represented "significant setbacks for civil rights."[59] 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

---

[55] Karen Yeung et al., *AI Governance by Human Rights Centered-Design, Deliberation and Oversight: An End to Ethics Washing, in* THE OXFORD HANDBOOK OF ETHICS OF AI 77, 77–105 (Markus D. Dubber et al. eds., 2020).

[56] Chris Jay Hoofnagle, *Assessing the Federal Trade Commission's Privacy Assessments*, 14 IEEE SEC. & PRIV. 58 (2016).

[57] *Id.* at 62.

[58] Laura W. Murphy, Final Report, *Facebook*'s *Civil Rights Audit*, META 1, 10 (July 08, 2020), https://about.fb.com/wp-content/uploads/2020/07/Civil-Rights-Audit-Final-Report.pdf.

[59] *Id.* at 8.

by the President and others who seek to use misinformation to sow confusion and suppress voting."[60] 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 the hiring of a senior official for civil rights advancement—but expressed concern that other decisions undermined progress. The auditors concluded that: "[u]nfortunately, 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."[61]

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 NPR that "[t]he 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."[62]

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,

> [t]he 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 consequential, 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

[60] *Id.* at 10.

[61] *Id.* at 8.

[62] Shannon Bond, *Report Slams Facebook for 'Vexing and Heartbreaking Decisions' on Free Speech*, NPR (July 08, 2020).

> dilemma would still come down to the decision of one person overseeing the communications of 1.73 billion people each day.[63]

The report also focused solely on the United States, at a time when Facebook's human rights record in non-U.S. 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.[64]

While Facebook clearly "failed" its civil rights audit, the meaning of failure must be questioned when the resulting recommendations were toothless. COO 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."[65] 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 fall 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.[66]

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

---

[63] Casey Newton & Zoe Schiffer, *What a damning civil rights audit missed about Facebook*, VERGE (July 10, 2020).

[64] Newley Purnell, *Facebook is Stifling Independent Report on its Impact in India, Human Rights Groups Say*, WALL ST. J. (Nov. 12, 2021).

[65] Sheryl Sandberg, *Making Progress on Civil Rights – But Still a Long Way to Go*, META (July 08, 2020), www.about.fb.com/news/2020/07/civil-rights-audit-report/.

[66] Laura W. Murphy, *The Rationale For and Key Elements of a Business Civil Rights Audit*, CIVIL RIGHTS DOCS (2021), www.civilrightsdocs.info/pdf/reports/Civil-Rights-Audit-Report-2021.pdf.

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.

II. ALGORITHMIC AUDITS IN LEGISLATION AND GOVERNMENTAL INQUIRIES

Legislation, proposed and 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 European Union, the United Kingdom, United States, and individual U.S. states.

A. *European Union*

The European Union'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.[67] The DSA also requires VLOPs that use recommendation systems to reveal in their Terms of Service the primary parameters used by algorithmic amplification systems.[68] Article 28 of the DSA requires VLOPs to submit yearly 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.[69]

---

[67] Luca Bertuzzi, *EU Institutions Reach Agreement on Digital Services Act*, EURACTIV (Apr. 23, 2022).

[68] James Vincent, *Google, Meta, and Others Will Have to Explain Their Algorithms under New EU Legislation*, VERGE (Apr. 23, 2022).

[69] Ilaria Buri & Joris van Hoboken, *The Digital Services Act (DSA) Proposal: A Critical Overview*, DSA OBSERVATORY 1, 37–38 (Oct. 28, 2021), https://dsa-observatory.eu/wp-content/uploads/2021/11/Buri-Van-Hoboken-DSA-discussion-paper-Version-28_10_21.pdf.

The final version provides some detail about auditor independence.[70] 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."[71] 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 systems.[72] 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-E.U. academics, researchers, and civil society groups also hope to be able to benefit from some of these transparency requirements.

Other E.U. 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.[73] The General Data Protection Regulation (GDPR) sets rules for the profiling of individuals and related automated decisionmaking and gives users the "right to explanation" about algorithmic processes.[74] 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 decisionmaking.[75] Commentators predict that this right, as well as entitlements to access

---

[70] *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*, COM (2020) 825 final (Jan. 20, 2022) (Article 28 provides that "[v]ery large online platforms shall ensure auditors have access to all relevant data necessary to perform the audit properly." Further, auditors must be "recognised and vetted by the Commission and . . . [must be] legally and financially independent from, and do not have conflicts of interest with" the audited platforms.).

[71] Buri & van Hoboken, *supra* note 69, at 37.

[72] Paddy Leerssen, *Platform Research Access in Article 31 of the Digital Services Act: Sword Without a Shield?* VERFASSUNGSBLOG (Sept. 07, 2021).

[73] *Platform-to-Business Trading Practices,* EUR. COMM'N, https://digital-strategy.ec.europa.eu/en/policies/platform-business-trading-practices (last visited Jan. 22, 2023); Věra Jourová, *The New Deal for Consumers: What Benefits Will I Get as a Consumer?*, EUR. COMM'N (Nov. 2019).

[74] GENERAL DATA PROTECTION REGULATION art. 22 (E.U.) (2018).

[75] Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 190, 206 (2019).

collected data, will lead to robust independent audits.[76] The E.U. 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.[77]

The E.U. draft Artificial Intelligence Act proposes a risk-based approach to AI regulation along a sliding scale of potential harms and requires in Article 61 that providers of high-risk AI systems conduct "conformity assessments" before their products enter the European market.[78] 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.[79] This approach to high-risk AI systems involves a combination of self-regulation, voluntary adherence to standards, and government oversight.[80]

B. *United States*

In the United States, a 2016 report by the Obama administration on algorithms and civil rights encouraged auditing.[81] Reintroduced in 2022, the Algorithmic Accountability Act would require the Federal Trade Commission (FTC) to create regulations and structures for companies to carry out assessments and provide transparency around the impact of automated decision-making.[82] 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

---

[76] Casey et al., *supra* note 23, at 150–51 (referring to the enhanced powers granted by the GDPR to E.U. data authorities).

[77] *Proposal For a Regulation of The European Parliament and of The Council on Contestable and Fair Markets in the Digital Sector (Digital Markets Act)*, COM (2020) 842 Final (Dec. 15, 2020).

[78] *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*, COM (2021) 206 final (Apr. 21, 2021).

[79] Natasha Lomas, *Europe's AI Act Contains Powers to Order AI Models Destroyed or Retrained, Says Legal Expert,* TECHCRUNCH (Apr. 01, 2022).

[80] Margot E. Kaminski, *Regulating the Risks of AI,* 103 B.U. L. REV. (forthcoming 2023) (manuscript at 51–54).

[81] *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights,* EXEC. OFF. PRESIDENT (May 2016), https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf.

[82] S. 3572, 117th Cong. (2022).

status."[83] While this seems like a step towards greater algorithmic fairness, it raises the question of what kind of fairness counts and how it should 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 others.[84] Moreover, fairness may conflict with other desirable goals of accuracy, efficiency, and privacy.[85]

The Digital Services Oversight and Safety Act,[86] introduced in 2022, would require the FTC to create regulations for large online platforms, requiring it to assess "systemic risks." These include 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."[87] The platforms would be required to commission an annual independent audit of their risk assessments and submit these to the FTC.[88] The American Data Privacy and Protection Act, released as a discussion draft in 2022, would require the developers of algorithms to collect, process, or transfer certain data necessary for the evaluation of algorithmic design (including any training data) in order to reduce the risk of civil rights harms.[89]

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

---

[83] *Id*.

[84] Richard N. Landers & Tara S. Behrend, *Auditing the AI Auditors: A Framework for Evaluating Fairness and Bias in High Stakes AI Predictive Models*, AM. PSYCH. 2–3 (2022) (discussing conceptions of fairness that range from transparency as to measurements and inputs, human rights, adverse effects on individuals, disadvantaging of underrepresented groups, and proprietary definitions of fairness and bias existing within corporations).

[85] Jess Whittlestone et al., *The Role And Limits Of Principles In AI Ethics: Towards A Focus On Tensions*, 19 CONF. ON AI, ETHICS, & SOC'Y 195 (2019) (discussing the gap between "principles and practical judgment" given that sometimes, "the benefit of using an algorithm may be high enough, and its accuracy reliable enough, that all users agree it is worth using even if a fully comprehensive explanation of its decisions cannot be given").

[86] H.R. 6796, 117th Cong. 2d Sess. (2022).

[87] *Id.* at 20.

[88] *Id.* at 44.

[89] H.R. 8152, 117th Cong. 2d Sess. (2022).

in algorithmic decision-making.[90] 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.[91]

The National Institute of Standards and Technology (NIST) published a draft risk management framework for AI systems in March 2022 recommending 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."[92] The NIST framework will ultimately be a guiding set of principles,[93] not binding legislation, and will not set substantive risk thresholds for companies.[94]

U.S. 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 that 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 systems that have not undergone the mandatory bias audit or for deployment of a system without disclosure to candidates.[95] In the limited context of pretrial risk assessment tools, the state of Idaho requires algorithmic transparency and open access to the public for the "inspection, auditing, and testing" of those tools.[96] Washington D.C.'s Attorney General has proposed a bill prohibiting algorithmic discrimination with respect to eligibility for "important life opportunities," and would require entities to audit their decisions and retain a five-year audit trail.[97]

Finally, the White House released a Blueprint for an AI Bill of

---

[90] S. 1896, 117th Cong. (2022).

[91] S. 3608, 117th Cong. (2022).

[92] *AI Risk Management Framework: Initial Draft*, NAT'L INST. STANDARDS & TECH. (NIST) 16 (Mar. 17, 2022), https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf.

[93] David Matthews, *How the US plans to manage artificial intelligence*, SCI.|BUS. (May 19, 2022).

[94] *AI Risk Management Framework: Initial Draft*, *supra* note 92.

[95] *AI Strategy: The New York City Artificial Intelligence Strategy*, N.Y.C. MAYOR'S OFF. CHIEF TECH. OFFICER (Oct. 2021), https://www1.nyc.gov/assets/cto/downloads/ai-strategy/nyc_ai_strategy.pdf; Automated Employment Decision Tools, Loc. L. 144 (N.Y.C. 2021).

[96] IDAHO CODE § 19-1902 (2022).

[97] D.C. CODE § B24-558 (2021–2022); Martin Austermuhle, *D.C. attorney general introduces bill to ban 'algorithmic discrimination*, NPR (Dec. 10, 2021).

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."[98] 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, in an interview with the Washington Post following its release.[99]

C. *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.[100] The Algorithmic Impact Assessment Tool questionnaire is a scorecard used to determine the impact level of an automated decision system.[101] 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.[102] Additional requirements are also imposed according to the impact level, which can include peer review, transparency, human intervention, contingency measures, or employee training.[103] Algorithmic impact assessments are mandatory for federal government institutions, with the exception of the Canada Revenue

---

[98] *Blueprint for an AI Bill of Rights*, WHITE HOUSE (Oct. 2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

[99] Cristiano Lima, *White House unveils 'AI Bill of Rights' as 'call to action' to rein in tool*, WASH. POST (Oct. 04, 2022).

[100] Treasury Board of Canada Secretariat, *Algorithmic Impact Assessment Tool,* GOV'T CAN. (last modified Jan. 19, 2023); Treasury Board of Canada Secretariat, *Directive on Automated Decision-Making*, GOV'T CAN. (Apr. 01, 2021).

[101] Treasury Board of Canada Secretariat, *Algorithmic Impact Assessment Tool*, *supra* note 100.

[102] Christine Ing et al., *Federal Government's Directive on Automated Decision-Making: Considerations and Recommendations*, MCCARTHY TETRAULT LLP (Apr. 13, 2019).

[103] Treasury Board of Canada Secretariat, *Directive on Automated Decision-Making*, *supra* note 100.

Agency.[104] 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.[105]

D. *Australia*

The 2021 News Media Bargaining Code governs commercial relationships between Australian news businesses and digital platforms,[106] requiring designated platforms to pay local news publishers for content linked on their platform and also requiring notice for changes to platform algorithms.[107] 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.[108] The proposal reads: "[d]esignated 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."[109]

E. *United Kingdom*

The draft U.K. Online Safety Bill gives regulator Ofcom significant investigatory power over platforms,[110] including the ability to audit algorithms of regulated entities.[111] Those entities must conduct risk assessments and then take steps to mitigate and manage identified

---

[104] Benoit Deshaies & Dawn Hall, *Responsible use of automated decision systems in federal government*, STAT. CAN. (Dec. 01, 2021).

[105] *Summary of Session Four: Regulatory Powers*, GOV'T CAN. (May 13, 2022); *Concluding Workshop Summary*, GOV'T CAN. (July 08, 2022).

[106] *News Media Bargaining Code,* AUSTL. COMPETITION & CONSUMER COMM'N (ACCC) (July 23, 2022).

[107] Asha Barbaschow, *Media Bargaining Code amendments include a more 'streamlined' algorithm change notice*, ZDNET (July 12, 2022).

[108] *Call For Tech Giants To Face Regular ACCC Algorithm Audits*, PARLIAMENT AUSTL. (Jan. 22, 2021).

[109] *Id.*

[110] *Draft Online Safety Bill*, DEPT. SCI., INNOVATION & TECH./DEPT. DIGIT., CULTURE, MEDIA & SPORT (May 2021), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/985033/Draft_Online_Safety_Bill_Bookmarked.pdf.

[111] *Findings from the DRCF Algorithmic Processing Workstream – Spring 2022*, INFO. COMM'RS OFF. (ICO) (Apr. 28, 2022).

risks of particular types of illegal and harmful content. Some service providers will also be required to publish transparency reports.[112] 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.[113] In addition, the Centre for Data Ethics and Innovation (CDEI), which is part of the Department for Digital, Culture, Media & Sport, has provided a Roadmap to an Effective AI Assurance Ecosystem.[114] 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."[115] The terms impact assessment, audit, and conformity assessment all show up in E.U. and U.K. legal instruments with particular meanings that are not the same as CDEI's.

[112] *Draft Online Safety Bill*, *supra* note 110.

[113] *Guidance on the AI auditing framework: Draft guidance for consultation*, ICO (July 23, 2022), https://ico.org.uk/media/2617219/guidance-on-the-ai-auditing-framework-draft-for-consultation.pdf.

[114] *The roadmap to an effective AI assurance ecosystem*, CTR. DATA ETHICS & INNOVATION (Dec. 08, 2021), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1039146/The_roadmap_to_an_effective_AI_assurance_ecosystem.pdf.

[115] *Id.*

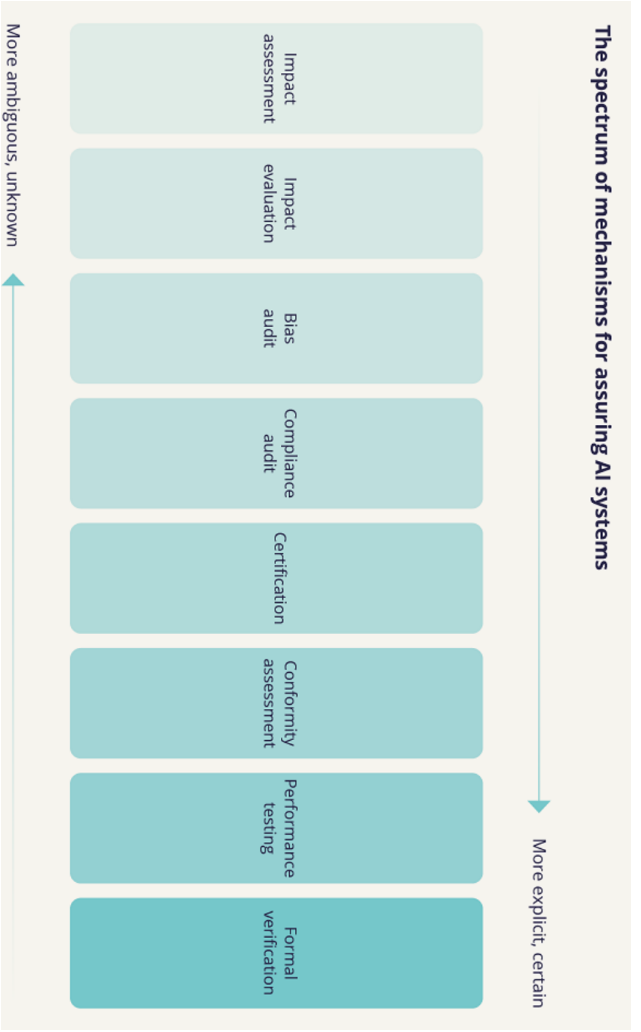

Figure 2 [116]

III. 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

[116] *Id.*

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.[117] Still, according to Twitter's Rumman Chowdhury, the field of reputable auditing firms is small—only ten to twenty.[118] 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.

A. *Who: Auditors*

Inioluwa Deborah Raji, a leading scholar of algorithmic audits, argues that the audit process should be interdisciplinary and multi-staged as it plays out, both internally for entities developing and deploying AI systems and externally for independent reviewers of those systems.[119]

Internal auditors, also known as first-party auditors, can intervene at any stage of the process.[120] Such auditors have full access to the system components pre-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 review, 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

---

[117] Kate Kaye, *A new wave of AI auditing startups wants to prove responsibility can be profitable*, PROTOCOL (Jan. 03, 2022).

[118] Alfred Ng, *Can Auditing Eliminate Bias from Algorithms?*, MARKUP (Feb. 23, 2021).

[119] Inioluwa Deborah Raji et al., *supra* note 23, at 37 (discussing "an initial internal audit framework" with "five distinct stages—Scoping, Mapping, Artifact Collection, Testing and Reflection (SMACTR)—all of which have their own set of documentation requirements and account for a different level of the analysis of a system"); Inioluwa Deborah Raji et al., *Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance*, PROC. 2022 AAAI/ACM CONF. ON AI, ETHICS, AND SOC'Y 560, 566 (June 09, 2022) ("AI policy can foster third party audits with more deliberate institutional design" centered around "five main considerations: (i) Target Identification & Audit Scope . . . (ii) Auditor Independence . . . (iii) Auditor Privileges . . . (iv) Auditor Professionalization & Conduct Standards . . . (v) Postaudit actions.").

[120] Inioluwa Deborah Raji et al., *supra* note 23, at 18.

audits that a company will conduct on its own.[121] 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 should be different from the team developing or using the tool subject to audit.[122] A number of commentators have called for increased rigor around internal auditing. Ifeoma Ajunwa, for example, proposes mandatory external auditing, in addition to internal auditing, for hiring algorithms.[123] 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."[124]

External audits necessarily look backwards and will be somewhat or entirely independent of 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.[125] 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.[126] For example, the company HireVue marketed its AI employment product as having passed a

---

[121] H.R. 6580, 117th Cong. 2d Sess. (Feb. 03, 2022).

[122] Comments from Andrew Burt and Solon Barocas in October 26, 2022 GMF-RIIPL workshop; *see* SR-11 Guidance on Model Risk Management, FED. RSRV. (Apr. 04, 2011); *Comptroller's Handbook: Model Risk Management*, OFF. COMPTROLLER CURRENCY (Aug. 2021), https://www.occ.treas.gov/publications-and-resources/publications/comptrollers-handbook/files/model-risk-management/pub-ch-model-risk.pdf.

[123] Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 HARV. J.L & TECH. 621, 659 (2021) (setting forth "mandated audits (both external and internal, which will enable litigation)" as one of several measures to hold automated hiring accountable).

[124] Shlomit Yanisky-Ravid & Sean K. Hallisey, *supra* note 23, at 429, 434.

[125] Raji et al., *Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance, supra* note 119.

[126] Mona Sloane, *The Algorithmic Auditing Trap*, ONEZERO (Mar. 17, 2021).

second-party civil rights audit, only for the independence of the auditors and the scope of the audit to be drawn into question.[127]

In order to ensure a degree of consistent rigor among auditors, Ben Wagner and co-authors have called for "auditing intermediaries."[128] 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.[129] 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.[130]

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-nineteenth century to protect shareholders from the improper actions of company directors.[131] 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."[132]

According to John Carey's history of American accounting practices, it was not until the turn of the twentieth century that financial

---

[127] Alex C. Engler, *Independent auditors are struggling to hold AI companies accountable*, FAST CO. (Jan. 26, 2021).

[128] Ben Wagner & Lubos Kuklis, *Establishing Auditing Intermediaries to Verify Platform Data*, *in* REGULATING BIG TECH: POLICY RESPONSES TO DIGITAL DOMINANCE (Martin Moore & Damian Tambini eds., 2021).

[129] *See, e.g.*, Ben Wagner et al., *Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act,* ACM CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY (2020) (reporting on audit findings that Facebook's reported data was wrong).

[130] *See* Ben Wagner & Lubos Kuklis, *supra* note 128; *see also* Ben Wagner et al., *The next step towards auditing intermediaries*, VERFASSUNGSBLOG (Feb. 23, 2022) (proposing that auditing intermediaries should serve a verification function (verifying accuracy of data provided by companies), facilitation function (lowering the costs of compliance for regulated companies and the costs of access for stakeholders), and a tailoring function (tailoring data access to individual needs thus reducing privacy risks)).

[131] Howard B. Levy, *History of the Auditing World, Part 1,* CPA J. (Nov. 2020).

[132] *Id.*

accountants started to organize and regulate themselves as a profession.[133] 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 followed the stock market crash of 1929: the Securities Act of 1933 and the 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[134]—which created a private nonprofit corporation to oversee audit procedures—and after the 2008 market crash with the 2010 Dodd-Frank Act[135]—which added to the requirements for independent audits and corporate audit committees, along with strengthening whistleblower protections.[136]

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.[137] 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 their systems.[138] External auditors will have to gain access

---

[133] John L. Carey, *Rise of the accounting profession, v. 1. From technician to professional, 1896-1936*, GUIDES, HANDBOOKS & MANUALS 30 (1969).

[134] H.R. 3763, 107th Cong. (2002).

[135] H.R. 4173, 111th Cong. (2009).

[136] Sarah J. Williams, *The Alchemy of Effective Auditor Regulation,* 25 LEWIS & CLARK L. REV. 1089, 1090, 1106, 1120–21 (2022) (The Sarbanes-Oxley Act of 2002 "significantly altered the formula for audit oversight by creating the PCAOB, a private, nonprofit corporation.").

[137] Jenna Burrell, *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, 3 BIG DATA & SOC'Y 1, 3 (2016) ("One argument in the emerging literature on the 'politics of algorithms' is that algorithmic opacity is a largely intentional form of self-protection by corporations intent on maintaining their trade secrets and competitive advantage.").

[138] Prerna Juneja et al., *Through the Looking Glass: Study of Transparency in Reddit's Moderation Practices*, 4 PROC. ACM ON HUM.-COMPUT. INTERACTION 1, 25 (2019) ("[S]tudy . . . revealed that several of the subreddits' rules are vaguely worded and their operationalization and enforcement is not transparent."); *Santa Clara Principles On Transparency and Accountability in Content Moderation*, SANTA CLARA PRINCIPLES, https://santaclaraprinciples.org (last visited Jul. 24, 2022).

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.[139]

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.[140] 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 U.S. Computer Fraud and Abuse Act if audited data is obtained without consent.[141] 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.[142] 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 examine the changes necessary to create legal certainty around algorithmic audits, both for internal audits by firms and external auditors or whistleblowers.[143]

---

[139] Hilke Schellmann, *Auditors Are Testing Hiring Algorithms for Bias, but There's No Easy Fix*, MIT TECH. REV. (Feb. 11, 2021).

[140] *See, e.g.*, Janne Chung et al., *Auditor Liability to Third Parties after Sarbanes-Oxley: An International Comparison of Regulatory and Legal Reforms*, 19 J. INT'L ACCT., AUDITING & TAX'N 66 (2010); Alan Reinstein et al., *Examining the Current Legal Environment Facing the Public Accounting Profession: Recommendations for a Consistent U.S. Policy*, 35 J. ACCT., AUDITING & FIN. 3, 3–25 (2020).

[141] 18 U.S.C. § 1030.

[142] *Sandvig v. Barr,* 451 F. Supp. 3d 73, 76 (D.D.C. 2020).

[143] Erwan Le Merrer et al., *Algorithmic Audits of Algorithms, and the Law*, HAL OPEN SCI. 1, 2 (2022) (urging that these questions be addressed for algorithmic audits: "i) what are the legal risks taken by an auditor and ii) can the outcome of an audit be used against its operating platform in court?").

B. *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."[144] 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 it can be unclear 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.[145] Critics responded that code alone does not make algorithms legible and accountable.[146] 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 the 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.[147] 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.

---

[144] IEEE, *IEEE Standard for Software Reviews and Audits*, IEEE COMPUT. SOC'Y 1, 30 (Aug. 15, 2008).

[145] Maxwell Adler, *Why Elon Musk Wants to 'Open Source' Twitter's Algorithms*, BLOOMBERG (Apr. 28, 2022).

[146] Cathy O'Neil, *Sorry Elon, 'Open Source' Algorithms Won't Improve Twitter*, WASH. POST (May 02, 2022).

[147] Deven R. Desai & Joshua A. Kroll, *Trust but Verify: A Guide to Algorithms and the Law*, 31 HARV. J.L. & TECH. 1, 10 (2017) (casting doubt on utility of transparency—"[f]or example, simply disclosing or open-sourcing source code does nothing to show that the disclosed software was used in any particular decision unless that decision can be perfectly replicated from the disclosures").

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.[148] 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."[149] 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.[150] So too, the AI Act's conformity assessments are to be done pre-deployment for high-risk systems per Articles 16 and 43.[151]

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.[152] 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 shows up, for example, in how the U.S. National Institute for Standards and Technology (NIST) conceptualizes the perpetuation of bias in AI:

---

[148] Andrew D. Selbst, *supra* note 38, at 146–47.

[149] Treasury Board of Canada Secretariat, *Algorithmic Impact Assessment Tool*, *supra* note 100.

[150] Dillon Reisman et al., *Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability*, AI NOW INST. (2018), https://openresearch.amsterdam/image/2018/6/12/aiareport2018.pdf.

[151] *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,* COM (2021) 206 Final (Apr. 21, 2021).

[152] Panel for the Future of Science and Technology (STOA), *Auditing the quality of datasets used in algorithmic decision-making systems*, EUR. PARL. 39 (July 2022), https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729541/EPRS_STU(2022)729541_EN.pdf.

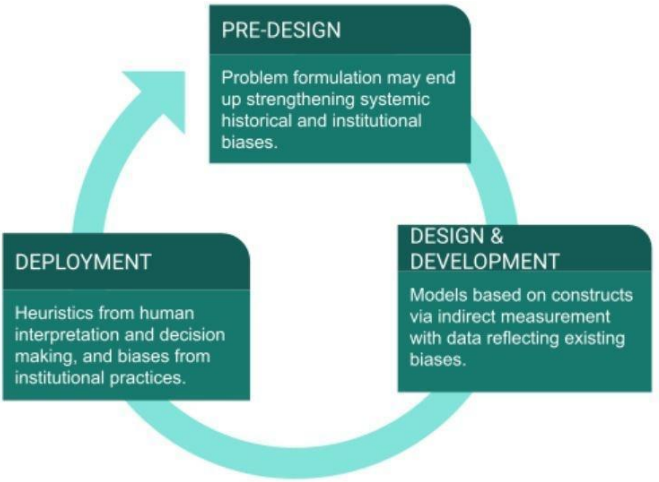

Figure 3 [153]

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."[154] 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.[155] 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

[153] Reva Schwartz et al., Draft NIST Special Publication 1270, *A Proposal for Identifying and Managing Bias in Artificial Intelligence*, U.S. DEPT. COM. NAT'L INST. STANDARDS & TECH., at 18 FIG. 2 (June 2021), https://doi.org/10.6028/NIST.SP.1270-draft.

[154] Miles Brundage et al., *Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims*, ARXIV 24 (Apr. 20, 2020).

[155] Mona Sloane et al., *A Silicon Valley love triangle: Hiring algorithms, pseudo-science, and the quest for auditability*, 3 PATTERNS 1, 3 (2022) ("If regulation and legislation are to be effective in producing accountability through mandating or recommending audit or assessment, methods for conducting such audits and assessments need to include new ways of framing and understanding how technological systems are encountered in the course of social life.").

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.[156] 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 blind spots when it comes to how "algorithmic monoculture" leads to this outcome homogenization.[157]

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 Cobbe and co-authors recommend a "holistic understanding of automated decision making 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."[158] Transparency around or audits of code alone will not be sufficient to reveal how algorithmic decision-making 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.[159]

---

[156] Andrew D. Selbst et al., *Fairness and Abstraction in Sociotechnical Systems*, PROC. CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 59 (2019) (arguing that definitions of fairness for machine learning systems "bound the system of interest [too] narrowly" by considering "the machine learning model, the inputs, and the outputs," and missing "the broader context, including information necessary to create fairer outcomes, or even to understand fairness as a concept").

[157] Comment by Deborah Raji in October 26, 2022 GMF-RIIPL workshop; *see Picking on the same person: Does Algorithmic Monoculture lead to Outcome Homogenization*, 36 CONF. NEURAL INFO. PROCESSING SYS. (Sept. 20, 2022) (Outcome homogenization is "the extent to which particular individuals or groups experience the same outcomes across different deployments.").

[158] Jennifer Cobbe et al., *supra* note 19.

[159] 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 06 2021); *Participatory Data Stewardship: A framework for involving people in the use of data*, ADA LOVELACE INST. (Sept. 07, 2021).

C. *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.[160] Alternatively, an audit may entail private internal 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 eighty-four AI ethics initiatives publishing frameworks.[161] Another fruitful source of objectives is the U.N. 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.[162] 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.[163] While research has shown that high-level ethical guidelines have not influenced the behavior of software engineers in the past,[164] 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

---

[160] For example, the consultation on a draft Canadian Online Safety bill talks about audit powers of a government entity to assess compliance with legal requirements. *Summary of Session Four: Regulatory Powers*, *supra* note 105 (discussing need for a Digital Safety Commissioner with audit powers). U.S. commentators have urged that the FTC audit commercial algorithms for deception and unfairness, meaning essentially that the agency investigate algorithmic functions as a predicate to legal action. *See, e.g.,* Danielle Keats Citron & Frank Pasquale*, The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV.1, 24–25 (2014) (proposing that the FTC audit consumer scoring systems).

[161] Brent Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI*, 1 NATURE MACH. INTELL. 501 (2019).

[162] *UN Guiding Principles Reporting Framework*, UNITED NATIONS, https://www.ungpreporting.org/ (last visited July 24, 2022).

[163] *Ethics Guidelines for Trustworthy AI,* EUR. COMM'N (Apr. 08, 2019).

[164] Thilo Hagendorff, *The ethics of AI ethics: An evaluation of guidelines*, 30 MINDS & MACHS. 99, 99–120 (2020).

system is biased against individuals or groups *vis-a-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.[165] 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.[166] Meta has now conducted a human rights audit itself,[167] but resisted submitting to external audits. Other inquiries relate to how platforms course-correct when new risks arise. The DSA and draft U.K. Online Safety Bill include auditing provisions for

[165] Alfred Ng, *supra* note 118; *Ethics Guidelines for Trustworthy AI, supra* note 163.

[166] *See* Aspen Digital, *Commission on Information Disorder Final Report*, ASPEN INST. (Nov. 16, 2021), www.aspeninstitute.org/publications/commission-on-information-disorder-final-report/; *see also* Rebecca Heilweil, *Facebook is taking a hard look at racial bias in its algorithms*, VOX (July 22, 2020, 1:12 PM).

[167] Miranda Sissons, *A Closer Look: Meta's First Annual Human Rights Report*, META (July 14, 2022).

mitigation.[168] 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.[169] 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 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.[170]

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."[171] Sasha Costanza-Chock and co-authors recommend that the applicable accountability framework be explicitly defined.[172] An audit that seeks

---

[168] *Draft Online Safety Bill, supra* note 110; Buri & van Hoboken, *supra* note 69, at 34.

[169] *Understanding Algorithms*, *supra* note 26, at 24.

[170] *An Audit of 9 Algorithms used by the Dutch Government*, NETH. CT. AUDIT (May 18, 2022).

[171] Mike Ananny & Kate Crawford, *supra* note 44, at 983.

[172] Sasha Costanza-Chock et al., *Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem*, 2022 ACM CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 1580 (2022) (recommending an accountability framework: "1) mandatory independent AI audits against clearly defined standards, applicable to both AI product owners and operators; 2) required notification to individuals when they are subject to algorithmic decision-making systems; 3) mandated disclosure of key components of audit findings for peer review; 4) consideration of real-world harm in the audit process, including standardized harm incident reporting and response mechanisms; 5) stakeholder participation in audits, in particular by communities most likely to experience

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 a consideration of potential attendant costs.[173] 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.[174] 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 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.

D. *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.[175] Some of this difficulty stems from the lack of agreed methods by which an audit is conducted. The question

---

harm from the system, product, or tool that is being audited; and 6) a formal system for evaluation and, potentially, accreditation of AI auditors").

[173] Inspired by observations by Niva Elkin-Koren, Professor of Law, Tel Aviv University, during GMF-RIIPL workshop, October 26, 2022.

[174] Jennifer Cobbe et al., *supra* note 19; Alene Rhea et al., *Resume Format, LinkedIn URLs and Other Unexpected Influences on AI Personality Prediction in Hiring: Results of an Audit,* PROC. 2022 AAAI/ACM CONF. ON AI, ETHICS & SOC'Y (2022).

[175] Briana Vecchione et al., *supra* note 8.

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.

U.K. 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.[176] The Ada Lovelace Institute has developed a taxonomy of social media audit methods by focusing on scraping, accessing data through APIs, and analyzing code.[177] 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, with 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 reliance on either custom tooling or mainstream resources that fall short of facilitating the actual audit goals of accountability."[178] 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 focusing too much on specific and technical notions of fairness or transparency that do not consider multiple stakeholders or the broader social context."[179] The U.K. Centre for Data Ethics and Innovation has announced that it will support the Department for Digital, Culture, Media and Sport (DCMS)'s Digital Standards team and the Office for AI (OAI) as they establish an AI Standards Hub, which focuses on global digital technical standards.[180] For the DSA, auditors like Deloitte propose applying their own methodologies:

> [t]he 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

[176] *Auditing Algorithms*, *supra* note 1.

[177] *Technical Methods for Regulatory Inspection of Algorithmic Systems*, ADA LOVELACE INST. (Dec. 09, 2021).

[178] Deb Raji, *It's Time to Develop the Tools We Need to Hold Algorithms Accountable*, MOZILLA (Feb. 02, 2022).

[179] Shea Brown et al., *supra* note 3, at 1.

[180] *The Roadmap to an Effective AI Assurance Ecosystem - Extended Version*, CTR. DATA ETHICS & INNOVATION (Dec. 08, 2021).

> required to meet these requirements ahead of the audit execution.[181]

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, U.S. 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.[182] The SEC has recognized standards adopted by the Financial Accounting Standards Board—a nonprofit consisting of a seven-person board—as authoritative.[183] In the tech context, a similar sort of co-regulation shapes Australia's Online Safety Act of 2021, the U.K. Online Safety Act, and the E.U. DSA, wherein all of which makes use of industry codes of conduct.[184] 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.

These codes might look like those being developed by the Partnership on AI, which is creating the codes of conduct for industry with respect to distinct problems like synthetic media and biometrics.[185] Still other standards will emerge from legacy standard-setting bodies, such as the IEEE, which has an initiative on Ethically Aligned Design.[186] In a 2019 report, this IEEE initiative said that "companies should make their systems auditable and should explore

---

[181] Mark Cankett & Lenka Fackovcova, *EU Digital Services Act: Are you ready for audit?*, DELOITTE (May 18, 2022).

[182] *See, e.g.*, Securities Act § 19(a) (1933) (codified at 15 U.S.C. § 77s(a)); Securities Exchange Act § 13(b)(1) (1934) (codified at 15 U.S.C. § 78m(b)(1)).

[183] *Policy Statement: Reaffirming the Status of the FASB as a Designated Private-Sector Standard Setter,* U.S. SEC. & EXCH. COMM'N (Apr. 25, 2003); *About the FASB*, FIN. ACCT. STANDARDS BD., https://www.fasb.org/info/facts (last visited July 23, 2022).

[184] *See generally* Julian Jaursch, *Overview of DSA delegated acts, reports and codes of conduct*, STIFTUNG NEUE VERANTWORTUNG (Aug. 15, 2022) (DSA Article 34(1)(d) explicitly mentions a potential "voluntary standard" for audits.).

[185] Claire Leibowicz, *PAI Developing Ethical Guidelines for Synthetic Media*, PARTNERSHIP ON AI (Mar. 10, 2022).

[186] *Ethically Aligned Design*, IEEE GLOB. INITIATIVE ON ETHICS AUTONOMOUS & INTELL. SYS., https://ethicsinaction.ieee.org/wp-content/uploads/ead1e.pdf (last visited July 23, 2022).

novel methods for external and internal auditing."[187] 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.[188] Alternatively, government entities themselves might set standards. For example, the E.U. 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.[189]

1. Case Study: Washington, D.C. "Stop Discrimination by Algorithms Act of 2021"

A proposed Washington D.C. regulation, "Stop Discrimination by Algorithms Act of 2021," would require algorithmic auditing by businesses making or supporting decisions on important life opportunities.[190] This 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 pretty clear, a 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 (Sec. 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 the D.C. Human Rights Act of 1977. There is also a transparency provision (Sec. 6), which requires covered entities to

---

[187] *Id.*

[188] Miles Brundage et al., *supra* note 154, at 3.

[189] High-Level Expert Group on Artificial Intelligence [hereinafter AI HLEG], *Ethics Guidelines for Trustworthy AI*, EUR. COMM'N (Apr. 08, 2019), https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf; AI HLEG, *Sectoral considerations on policy and investment recommendations for trustworthy AI,* EUR. COMM'N 3 (July 23, 2020).

[190] Stop Discrimination by Algorithms Act of 2021, B. 24-558, 24th Council, Reg. Sess. (D.C. 2021).

provide notice of their use of personal information in algorithmic practices and notices and explanations of adverse decisions.

The audit provision (Sec. 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 D.C. 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.

2. 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.[191] 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 1000 children in foster care based on "risk factors" like dual nationality or low income.[192]

The audit was based on a framework laid out in the 2021 report "Understanding Algorithms" from the Netherlands Court of Audit.[193] The auditing framework is publicly available for download.[194] The framework assesses algorithms across five metrics: governance and accountability; model and data; privacy; IT general controls; and

---

[191] *An Audit of 9 Algorithms used by the Dutch Government*, *supra* note 170.

[192] Melissa Heikkila, *Dutch scandal serves as a warning for Europe over risks of using algorithms*, POLITICO (Mar. 29, 2022).

[193] *Understanding Algorithms*, *supra* note 26.

[194] *Audit Framework for Algorithms*, NETH. CT. AUDIT (Jan. 26, 2021).

ethics, which encompasses respect for human autonomy; the prevention of damage; fairness; explicability; and transparency.

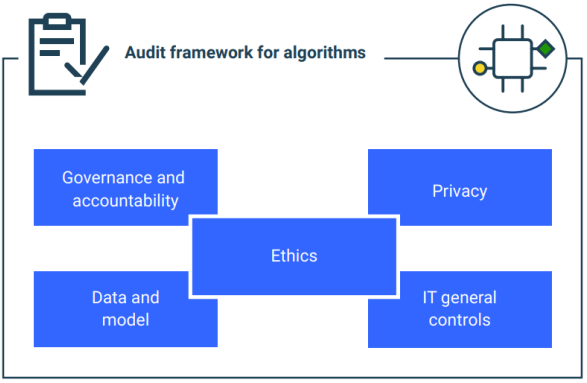

Figure 4 [195]

No.	Ethical framework	Ethical principle
1.1	Respect for human autonomy	The decisions made by the algorithm are open to human checks.
2.1	The prevention of damage	The algorithm is safe and always does what it is supposed to do.
2.2		Privacy is safeguarded and data protected.
3.1	Fairness	The algorithm takes account of diversity in the population and does not discriminate.
3.2		The algorithm's impact on society and the environment was taken into account during its development.
4.1	Explicability and transparency	It is possible to explain the procedures that have been followed.
4.2		It is possible to explain how the algorithm works.

Figure 5 [196]

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

[195] *Id.* at 14.
[196] *Id.* at 48.

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?[197]

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.[198]

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 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

[197] *An Audit of 9 Algorithms used by the Dutch Government*, *supra* note 170, at 42.

[198] *Id.* at 43.

national regulator to ensure audit quality and necessary changes if an algorithm or firm fails.

IV. CONCLUSION

Audits of automated decision systems, also called algorithmic or AI systems, are currently required in some cases by the E.U. Digital Services Act, arguably by the E.U. GDPR, and either required or considered in a host of U.S. 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 their proponents imagine, our review of the relevant provisions and proposals the term audit and associated terms require much more precision.

1. *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.

2. *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.

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

4. *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.