USER-BASED ALGORITHMIC AUDITING

URI Y. HACOHEN[†]

In the artificial intelligence and cloud computing age, digital platforms like Meta, Google, and Amazon wield immense social, economic, and political power, shaping users' daily lives. As these platforms gather vast amounts of user data and utilize sophisticated algorithms to personalize services, they also expose users to risks of bias and manipulation. Policymakers seek ways to hold platforms accountable, and algorithmic auditing is emerging as a key approach. However, existing regulations often rely on self-audits by the platforms themselves, leading to conflicts of interest. The shift towards third-party auditing is promising but still falls short of resolving these conflicts. To address this challenge, this article introduces, typologizes and explores a unique and underutilized approach to regulatory algorithmic oversight: "user-based algorithmic auditing." According to this auditing approach, the platforms' users lead the audit or assist external auditors in the process. User-based auditing is impartial, as it is entirely independent of the audited platforms. User-based audits are also valuable for corroborating the information the platforms provide in their self-auditing reports. The article explores regulatory frameworks, scrutinizes auditing approaches, and delves into the potential of user-based auditing to shape algorithmic oversight policies effectively.

[†] Professor of Law and Faculty Fellow at the Chief Justice Meir Shamgar Center for Digital Law and Innovation, The Buchmann Faculty of Law, Tel Aviv Uni-versity. I would like to thank Professor Niva Elkin-Koren and Professor Anat Ben-David for the helpful comments. Many thanks also for Yonati Bonn Yavneh and Roni Berman for wonderful research assistance and insightful conversations.

I. Introduction

Driven by the economics of scale and scope associated with artificial intelligence (AI) and cloud computing, digital platforms like Meta, Google, and Amazon amass tremendous social, economic, and political power.[1] These platforms increasingly impact their users' daily lives, from facilitating political conversations to warming children's bedrooms on a chilly night.[2] Users maintain close relationships with these platforms. The platforms provide valuable services to users while simultaneously tracking and profiling the users' behavior.[3] Platforms collect, store, and analyze massive amounts of user data and employ sophisticated machine-learning algorithms to extract insights from that data to optimize and personalize their services.[4] In this way, platforms direct their users' engagement with services, other users, and businesses.[5]

Users gain substantial value from the platforms' services but also expose themselves to systematic risks of algorithmic bias, discrimination, and manipulation.[6] With their growing power, society expects digital platforms to assume growing responsibility.[7] Indeed, policymakers are debating different ways to hold these platforms accountable through regulation, taxes, and even structural corporate divestiture.[8] As a complementary measure, algorithmic auditing emerges as the least controversial approach.[9]

Algorithmic auditing is an umbrella term used for attempts by various stakeholders to scrutinize the platforms' "black box" algorithms for social harm.[10] Algorithmic audits have successfully

[1] Uri Y. Hacohen, *Policy Implications of User-Generated Data Network Effects*, 33 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 340 (2023) [Hacohen, Policy]; Uri Y. Hacohen, *User-Generated Data Network Effects and Market Competition Dynamics*, (Forthcoming in FORDHAM INTELL. PROP. MEDIA & ENT., 2023) [Hacohen, Competition].

[2] Brian Heater, *Alexa will warm Eight's new Smart Mattress*, TECHCRUNCH (Nov. 22, 2016).

[3] *See* Hacohen, Competition, *supra* note 1.

[4] *Id.*

[5] *See e.g.,* Binh Le, Damiano Spina, Falk Scholer & Hui Chia, *A Crowdsourcing Methodology to Measure Algorithmic Bias in Black-Box Systems: A Case Study with COVID-Related Searches*, *in* ADVANCES IN BIAS AND FAIRNESS IN INFORMATION RETRIEVAL 43 (Ludovico Boratto, Stefano Faralli, Mirko Marras & Giovanni Stilo eds., 2022).

[6] *See, e.g.,* Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACHINE. LEARNING RES. 77, 78 (2018) (discussing biases in gender classification systems); U.S. GOV'T ACCOUNTABILITY OFF., FACIAL RECOGNITION TECHNOLOGY 1, 24–26 (July 2020); Allison Koenecke, et al., *Racial Disparities in Automated Speech Recognition*, 117 PROC. NAT'L ACAD. SCI. 7684, 7684 (2020) (identifying biases speech recognition applications); Jieyu Zhao et al., *Men Also Like Shopping: Reducing Gender Bias Amplification Using Corpus-Level Constraints*, in PROCEEDINGS OF THE 2017 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING 2979–2989 (2017), https://aclanthology.org/D17-1323 (discussing biases in image search). *See generally* Paresh Dave, *Fearful of Bias, Google Blocks Gender-Based Pronouns from New AI Tool*, REUTERS (Nov. 26, 2018) https://www.reuters.com/article/usalphabet-google-ai-gender/fearful-of-bias-google-blocks-gender-based-pronouns-fromnew-ai-tool-idUSKCN1NW0EF (describing biases in predictive text). *See generally* Hacohen, Competition, *supra* note 1; Bar-Gil, *supra* note 7.

[7] See e.g., Oren Bar-Gill, Cass R. Sunstein, & Inbal Talgam-Cohen, *Algorithmic Harm in Consumer Markets* (Preliminary draft, 2022).

[8] Hacohen, *supra* note 1.

[9] *See generally* Digital Regulation Cooperation Forum [hereinafter DRCF], *Auditing Algorithms: the Existing Landscape, Role of Regulators and Future Outlook*, GOV.UK (Sept. 23, 2022), https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook.

[10] *See generally*, Colleen Honigsberg et al., *Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance*, ARTIFICIAL INTELLIGENCE, ETHICS, & SOC'Y (2022) https://arxiv.org/pdf/2206.04737.pdf.

uncovered failures and biases in automated systems in areas such as housing,[11] employment markets,[12] web search,[13] social media,[14] and e-commerce.[15]

All major jurisdictions are currently devising new algorithmic auditing regulations.[16] However, in many preexisting schemes, the auditing entities are (somewhat paradoxically) the audited platforms themselves. For example, the EU's General Data Protection Regulation (GDPR) and the proposed US Algorithmic Accountability Act require the platforms themselves to generate "risk assessments" of their own algorithms.[17] These so-called "first-party" auditing approaches are problematic.[18] Platforms are commercial, profit-driven entities whose private interests often fail to align with the societal interests that algorithmic audits seek to safeguard. To illustrate this tension, just consider Google's recent dismissal of researchers criticizing the company's language models[19] or Facebook's alleged attempt to cover up evidence of misinformation going viral on its platform.[20]

To remedy this conflict of interests, policymaking is shifting towards "third-party" audits, where the auditing entities—governments, academics, or journalists—are independent of the platforms they audit.[21] For example, the European Digital Services Act (DSA) requires platforms to provide information to academic researchers who serve as external auditors.[22] Similarly, the Algorithmic Accountability Act and the Platform Accountability and Transparency Act require platforms to share information with the Federal Trade Commission (FTC).[23]

The move from first-party to third-party audits is welcome but still insufficient. Under most preexisting regulations, the information underlying third-party review originates from and is fully

[11] *See e.g.,* Joshua Asplund, Motahhare Eslami, Hari Sundaram, Christian Sandvig, and Karrie Karahalios, *Auditing Race and Gender Discrimination in Online Housing Markets,* PROCEEDINGS OF THE INTERNATIONAL AAAI CONFERENCE ONWEB AND SOCIAL MEDIA 14, 1 (May 2020), 24–35. https://ojs.aaai.org/index.php/ICWSM/article/view/7276

[12] *See e.g.,* Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson, *Investigating the Impact of Gender on Rank in Resume Search Engines. Association for Computing Machinery*, NEW YORK, NY, USA, 1–14 (2018) https://doi.org/10.1145/3173574.3174225 ; Jeffrey Dastin, *Amazon scraps secret AI recruiting tool that showed bias against women*. (2018). https://www.reuters.com/article/us-amazon-com-jobsautomation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showedbias-against-women-idUSKCN1MK08G

[13] *See e.g.,* Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism.* (2018) https://www.jstor.org/stable/j.ctt1pwt9w5; Ronald E. Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, & Christo Wilson, *Auditing Partisan Audience Bias within Google Search. Proc.* ACM HUM.-COMPUT. INTERACT. 2, CSCW, 148 (Nov. 2018) https://doi.org/10.1145/3274417

[14] *See e.g.,* Juhi Kulshrestha, Motahhare Eslami, Johnnatan Messias, Muhammad Bilal Zafar, Saptarshi Ghosh, Krishna P. Gummadi, & Karrie Karahalios, *Quantifying Search Bias: Investigating Sources of Bias for Political Searches in Social Media*, CORR ABS/1704.01347 (2017). arXiv:1704.01347 http://arxiv.org/abs/1704.01347

[15] *See e.g.,* Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson, *Measuring Price Discrimination and Steering on E-Commerce Web Sites* (2014), 305–318. https://doi.org/10.1145/2663716.2663744

[16] *See infra* Part II.

[17] *Id.*, at 3.

[18] *Id.*, at 3.

[19] Cade Metz & Daisuke Wakabayashi, *Google Researcher Says She Was Fired Over Paper Highlighting Bias in A.I.*, N.Y. TIMES (Dec. 3, 2020).

[20] Jeff Horwitz, *The Facebook Whistleblower, Frances Haugen, Says She Wants to Fix the Company, Not Harm It*, WALL ST. J. (Oct. 3, 2021), https://www.wsj.com/articles/facebook-whistleblower-frances-haugen-says-she-wants-tofix-the-company-not-harm-it-11633304122;

[21] Honigsberg, *supra* note 10.

[22] Adam Satariano, *E.U. Takes Aim at Social Media's Harms with Landmark New Law*, N.Y. TIMES (Apr. 22, 2022) https://www.nytimes.com/2022/04/22/technology/europeanunion-social-media-law.html; Frances Haugen, *Europe Is Making Social Media Better Without Curtailing Free Speech. The U.S. Should, Too*, N.Y. TIMES (Apr. 28, 2022), https://www.nytimes.com/2022/04/28/opinion/social-mediafacebook-transparency.html ("The new requirement for access to data will allow independent research into the impact of social media products on public health and welfare.").

[23] Ben Smith, *A Former Facebook Executive Pushes to Open Social Media's 'Black Boxes,'* N.Y. TIMES, (Jan. 2, 2022), https://www.nytimes.com/2022/01/02/business/media/crowdtangle-facebook-brandon-silverman.html; American Data Privacy and Protection Act, H.R. 8152, § 202 117th Cong. (2022); The Algorithmic Accountability Act of 2022, H.R. 6580, § 6 117th Cong. (2022).

controlled by the audited platforms. As such, third-party audits cannot remedy the inherent conflict of interests associated with self-auditing.[24] Worse, in some cases, third-party auditing may even intensify, rather than resolve, the algorithmic harms by affording the platforms an illusory seal of regulatory approval.[25]

To deal with these emerging problems, this article typologizes and explores a unique and underutilized approach to regulatory algorithmic oversight: "user-based algorithmic auditing."[26] According to this auditing approach, users initiate or govern the algorithmic auditing process or assist third-party auditors (such as governments or academics) in performing the task.[27] User-based audits are more impartial than other types of third-party audits because they do not depend on the audited platforms for information. User-based audits can also corroborate the information that platforms provide in first-party (or commissioned "second-party") audits.[28] For example, auditors can corroborate the authenticity of platforms' official content-removal policies by comparing them with real-world user data about content removal practices.[29]

This article has three parts. Part II explores the preexisting regulatory frameworks for algorithmic auditing in the US, EU, and UK. Part III introduces and examines the traditional classification of algorithmic audits. Lastly, Part IV presents a typology of user-based algorithmic auditing, explores its implications, and offers an array of policies to make it more salient. The proposals suggested in this article have the potential to improve algorithmic oversight and enable platform research and governance.

II. Algorithmic Auditing Regulations

As with other emerging technologies, the lack of institutional competency and the slow pace of regulatory decision-making make algorithmic auditing regulation a challenging endeavor. These challenges are especially severe in algorithmic auditing because machine-learning and AI are complex and opaque technologies with an ever-increasing range of applications.[30]

Given these challenges, it is extremely difficult to devise regulations that will adequately address algorithmic harms. Worse, premature or ill-devised regulations may prove counterproductive by creating a false aura of supervisory assurance.[31] Fears of such "regulatory washing" were invoked in the privacy context. For instance, when the FTC settles allegations of privacy violations, the agency often asks the settling companies to obtain outside assessments of the firm's privacy and security programs.[32] Relying on the FTC's reputation, the general public may find these settlements assuring.

[24] *See infra* Part III.
[25] *See infra* Part III.
[26] *See infra* Part IV.
[27] *See infra* Part IV.
[28] While generic whistleblower statutes such as the False Claims Act and the Whistleblower Protection Act will be applicable to the public trust in UGD, additional sui generis whistleblower provisions might be needed. For example, users could be rewarded by flagging opaque content removals or suspected personalized suggestions. Regulators could compare the data provided by the users to the data provided by the data platforms to detect systematic biases or inaccuracies in their algorithmic audits.
[29] *See infra* Part IV.
[30] *See generally* DAVID FREEMAN ENGSTROM ET AL, GOVERNMENT BY ALGORITHM: ARTIFICIAL INTELLIGENCE IN FEDERAL ADMINISTRATIVE AGENCIES (REPORT SUBMITTED TO THE ADMINISTRATIVE CONFERENCE OF THE US, (Feb. 2020).
[31] Goodman & Trehu, *supra note* 64.
[32] Chris Jay Hoofnagle, *Assessing the Federal Trade Commission's Privacy Assessments*, 14(2) IEEE SECURITY & PRIVACY 58-64 (2016).

Nevertheless, researchers have shown that such settlements sometimes conform to the companies' agenda, not to independent third-party standards.[33]

Algorithmic auditing regulations may also lead to severe regulatory conflicts.[34] This problem is significant, not only because different jurisdictions often disagree on which social values algorithmic systems should prioritize, but also because the mere definitions of the values themselves vary.[35] For example, the computer science literature offers over twenty definitions for "fairness."[36]

Despite these concerns, nearly all major Western jurisdictions are devising algorithmic auditing regulations. This section briefly summarizes the emerging regulatory frameworks in three major jurisdictions: the United States, the European Union, and the United Kingdom.

A. United States

A surge of new algorithmic auditing regulations is pending consideration by the US Congress.[37] First introduced on February 3, 2022, the Algorithmic Accountability Act requires platforms to perform bias impact assessments of their automated decision-making systems.[38] Similarly, the 2022 Digital Services Oversight and Safety Act would require platforms to assess "systemic risks" of illegal content and violation of community standards.[39] The Act would require platforms to report these risks to the FTC annually.[40]

Another ambitious legislation, the Platform Accountability and Transparency Act, would require platforms to explain how their recommendation and ranking algorithms work and to provide statistics on their content moderation actions.[41] The Act would also require platforms to share data with journalists and researchers in a way that could facilitate third-party auditing by these entities.[42] A few

[33] Goodman & Trehu, *supra note* 64, at 10, Hoofnagle, *id.*

[34] The problem of regulatory conflict is already manifesting in privacy regulation, triggering an unofficial *de facto* standardization process that is sometimes labeled "the Brussels effect." Anu Bradford, The Brussels Effect: How The European Union Rules The World, 149 (2020), https://oxford.universitypressscholarship.com/view/10.1093/oso/9780190088583.001.000 1/oso-9780190088583-chapter-6 [https://perma.cc/8B39-C5E7] ("[C]ompanies . . . fear[] the emergence of a complex patchwork of potentially conflicting state privacy laws.").

[35] Goodman & Trehu, *supra note* 64.

[36] Doaa Abu Elyounes, *Contextual Fairness: A Legal & Policy Analysis of Algorithmic Fairness*, U. Ill. JL. Tech. & Pol'y, 1 (2020), ahil Verma & Julia Rubin, *Fairness Definitions Explained, FairWare* '18 Proceedings of the International Workshop on Software Fairness 1, 2–3 (2018).

[37] *See e.g.,* Makenzie Holland, Experts call for AI regulation during Senate hearing | TechTarget (Mar. 10, 2023) https://www.techtarget.com/searchcio/news/365532338/Experts-call-for-AI-regulation-during-Senate-hearing (last visited Jul 19, 2023). Most recently, the National Telecommunications and Information Administration has issued an AI Accountability Policy Request for Comments. National Telecommunications and Information Administration, *AI Accountability Policy Request for Comment*, Federal Register (2023), No. 230407–0093 (Apr. 13, 2023), https://www.federalregister.gov/documents/2023/04/13/2023-07776/ai-accountability-policy-request-for-comment (last visited Jul 19, 2023). The agency has received comments from various stakeholds and is currently contemplating policies for regulatory action.

[38] Gibson Dunn, *Artificial Intelligence and Automated Systems Legal Update (1Q22)*, Gibson Dunn (May 5th, 2022), https://www.gibsondunn.com/artificial-intelligence-and-automated-systems-legal-update-1q22/.

[39] Goodman & Trehu, *supra note* 64. at 15.

[40] Goodman & Trehu, *supra note* 64. at 15.

[41] Platform Accountability and Transparency Act (Draft Bill), S. ____, 117th Cong., https://perma.cc/8C7Z-NSMN. *See also* John Perrino, *Platform Accountability and Transparency Act Reintroduced in Senate*, Stanford Cyber Policy Center (Jun. 8, 2023), https://cyber.fsi.stanford.edu/news/platform-accountability-and-transparency-act-reintroduced-senate (last visited Jul 19, 2023).

[42] *Id.*

other proposed Bills and pending policies suggest imposing transparency and data-sharing obligations on platforms.[43]

Some US states and even municipalities are also considering their own algorithmic auditing regulations. For example, New York City has recently published a detailed AI strategy.[44] As part of this strategy, the city has recently passed legislation requiring companies to use AI-based hiring tools to commission independent bias audits and to disclose to their applicants how their algorithmic systems are being used.[45] Similarly, Washington DC's Attorney General has proposed a Bill prohibiting algorithmic discrimination regarding "important life opportunities."[46] According to this Bill, companies must retain a five-year audit trail for algorithmic decisions.[47]

B. European Union

The landmark 2022 Digital Services Act (DSA) requires "very large online platforms" (VLOPs) to conduct annual systemic risk assessments of the online harms they create and to mitigate these harms.[48] The DSA also proposes a mechanism for facilitating data access to vetted researchers and other external auditors to explore the platforms' algorithmic systems.[49] However, because the Act defines "vetted researchers" rather loosely, it is currently unclear whether stakeholders such as non-academic researchers and civil society groups could benefit from the Act's transparency requirements.[50]

Another key EU regulation that has algorithmic auditing responsibilities is the 2016 General Data Protection Regulation (GDPR). The GDPR obligates platforms to conduct 'Data Protection Internal Audits,' (DPIAs) and employ a data protection officer.[51] The Act also requires platforms to share information with regulatory authorities[52] and empowers these authorities to perform algorithmic audits.[53]

[43] *See. E.g.,* Justin Hendrix, *Senators Blackburn and Blumenthal Unveil Kids Online Safety Act*, TECH POLICY PRESS (Feb. 16, 2022), https://techpolicy.press/senators-blackburn-and-blumenthal-unveil-kids-online-safety-act/ (last visited Jul 19, 2023) (Introducing The Kids Online Safety Act which includes a component for researcher access to study social media sites for specific harms; The White House, *U.S.-EU Joint Statement of the Trade and Technology Council*, THE WHITE HOUSE (May. 31, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/05/31/u-s-eu-joint-statement-of-the-trade-and-technology-council-2/ (last visited Jul 19, 2023) (A joint statement from the White House and European Commission following the summit said that "it is crucially important for independent research teams to be able to investigate, analyze and report on how online platforms operate and how they affect individuals and society.").

[44] Nicol Turner Lee & Samantha Lai, *Why New York City is cracking down on AI in hiring*, BROOKINGS (Dec. 20, 2021)

[45] *Id.*

[46] Goodman, *supra note 6*.

[47] *Id.*

[48] Articles 26, 27 the Digital Service Act; *See* Ilaria Buri & Joris van Hoboken, *The Digital Services Act (DSA) proposal: a critical overview, Institute for Information Law (IViR)*, University of Amsterdam, Discussion paper (Oct. 2021), https://dsa-observatory.eu/wp-content/uploads/2021/11/Buri-Van-Hoboken-DSA-discussion-paper-Version-28_10_21.pdf; Luca Bertuzzi, *EU Institutions Reach Agreement on Digital Services Act,*" EURACTIV, April 23, 2022, https://www.euractiv.com/section/digital/news/eu-institutions-reach-agreement-on-digital-services-act/.

[49] Article 31 the Digital Service Act; Paddy Leerssen, *Platform Research Access in Article 31 of the Digital Services Act: Sword without a Shield?* VERFASSUNGSBLOG, September 7, 2021, https://verfassungsblog.de/power-dsa-dma-14/.

[50] Goodman & Trehu, *supra note* 64. at 14.

[51] Art. 35 GDPR.

[52] Art. 58(1)(e) GDPR.

[53] Art. 58(1)(b) GDPR. In addition, when the platforms' algorithmic systems may affect specific individuals or human rights, then the platforms' algorithmic systems must be audited by an independent third-party entity. *See* PAWEL KUCH, TAMING THE ALGORITHM: THE

Other related regulations are the Platform-to-Business Regulation 2020 and the New Deal for Consumers form 2017, which require platforms to disclose the general parameters of their algorithmic ranking systems to businesses and consumers, respectively.[54] Lastly, the EU's AI Act 2021 mandates conformity assessment for high-risk applications.[55] Platforms may carry out typical audits internally, but for high-risk cases, the Act requires external audits by notified bodies.[56]

C. United Kingdom

The UK is highly invested in algorithmic auditing regulation. To facilitate such regulation, the government has created the Digital Regulation Cooperation Forum (DRCF), a designated platform for four government agencies to collaborate in their regulatory efforts.[57] The DRCF includes the Competition and Market Authority (CMA), the Financial Conduct Authority (FCA), the Information Commissioner's Office (ICO), and the Office of Communications (Ofcom).[58]

The DRCF authority comes from several laws, including The Competition Act (1998), the Consumer Rights Act (2015), the Financial Services and Markets Act (2000), the UK General Data Protection Regulation, and the Data Protection Act (2018).[59] The forum members have broad legal powers, such as gathering information (including data and code) from the audited platforms and conducting on-site tests and interviews.[60]

Recent announcements and publications indicate that the UK's interest in algorithmic auditing regulation will grow in the incoming years. These publications include the Centre for Data Ethics and Innovation's AI Assurance roadmap, the pilot for an AI Standards Hub for coordinating UK engagement in AI standardization globally, and the Central Digital and Data Office's Algorithmic Transparency Standard, which fosters algorithmic transparency in public sector organizations.[61] In addition, the UK's National Audit Office is collaborating with public audit organizations in Norway, the Netherlands, Finland, and Germany to devise a list of contemporary best practices for algorithmic auditing regulations.[62]

RIGHT NOT TO BE SUBJECT TO AN AUTOMATED DECISION IN THE GENERAL DATA PROTECTION REGULATION 137 (2022), https://eizpublishing.ch/wp-content/uploads/2022/09/Taming-the-algorithm-Digital-V1_00-20220826.pdf

[54] Goodman & Trehu, *supra note* 64 at 14; *Platform-to-Business Trading Practices*, EUROPEAN COMMISSION, June 7, 2022, https://digitalstrategy.ec.europa.eu/en/policies/platform-businesstradingpractices#:~:text=The%20EU%20Regulation%20on%20platform,and%20traders%20on%20online%20platforms; Věra Jourová, *The New Deal for Consumers: What Benefits Will I Get as a Consumer?* EUROPEAN COMMISSION, November 2019, https://ec.europa.eu/info/sites/default/files/consumer_agenda_-_factsheet_-_en.pdf.

[55] AUDITING ALGORITHMS: THE EXISTING LANDSCAPE, ROLE OF REGULATORS AND FUTURE OUTLOOK, DIGITAL REGULATION COOPERATION FORUM (DRCF), April 28, 2022, https://www.gov.uk/government/publications/findings-from-the-drcfalgorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulatorsand-future-outlook [hereinafter: DRCF].

[56] DRCF, *id.*

[57] DRCF, *id.*

[58] DRCF, *id.*

[59] DRCF, *id.*

[60] DRCF, *id.*

[61] DRCF, *id.*

[62] DRCF, *id.*

III. Types of Algorithmic Audits

"Audit" is an umbrella term that describes efforts to research and evaluate complex processes to determine whether they comply with predefined policies or regulations.[63] Algorithmic audits use technical and non-technical measures to detect and explain systematic errors in algorithmic systems.[64] The primary purpose of these processes is to align algorithmic systems employed by private profit-driven businesses with social values such as explainability, transparency, fairness, privacy, due process, accountability, robustness, and security.[65]

While the substantive requirements for conducting algorithmic audits may vary with the risks involved, all auditing regimes can be classified according to some essential characteristics.[66] One standard classification of algorithmic auditing regimes is based on the relationship between the auditing entity and the audited company.[67] This classification distinguishes between *first-party* auditors (the audited platforms themselves), *second-party* auditors (external auditors that collaborate with the audited platforms), and *third-party* auditors (auditors that are external and independent from the audited company).[68]

As depicted in Figure 1, there is a clear inverse relationship between the auditors' technical capability to conduct proper algorithmic audits and their motivation to pursue this task. First-party auditors have the technical expertise and uninterrupted access to algorithms and to users' data, which allows them to execute algorithmic audits in real-time.[69] However, as profit-driven businesses, first-party auditors are naturally motivated to maximize shareholder value, not to align algorithmic performance with social values.[70] Conversely, third-party auditors are the most motivated but the least capable of conducting algorithmic audits. While their values may align with the social values the audits seek to promote, third-party auditors lack the experience and the access to algorithms and data that first-party auditors have. Second-party auditors are situated between these extremes. The following sections explore these dynamics in greater detail.

[63] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron & Parker Barnes, *Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic*, FAT* 33 (2020) [hereinafter *Closing the Gap*]; DRCF, *supra* note 9.

[64] J. Nathan Matias, Austin Hounsel & Nick Feamster, *Software-Supported Audits of Decision-Making Systems: Testing Google and Facebook's Political Advertising Policies*, CS.HC. 1 (2021); Ellen P. Goodman & Julia Trehu, *AI Audit-Washing and Accountability*, SSRN (Sep. 30, 2022), https://ssrn.com/abstract=4227350 [hereinafter Goodman].

[65] Goodman & Trehu, *supra* note 64.

[66] Ellen Goodman Julia Trehu, for example, classify audits based on party who conducts the audit, the subject matter audited, the purpose of auditing, and the nature of the auditing process. Goodman & Trehu, *supra note* 64. *See also* ADA LOVELACE INSTITUTE, TECHNICAL METHODS FOR REGULATORY INSPECTION OF ALGORITHMIC SYSTEMS, (Dec. 9, 2021) (proposing a taxonomy of social media audit methods, focusing on scraping, accessing data through application programming interfaces, and analyzing code).

[67] *See generally*, Colleen Honigsberg et al., *Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance*, ARTIFICIAL INTELLIGENCE, ETHICS, & SOC'Y (2022) https://arxiv.org/pdf/2206.04737.pdf.

[68] DRCF, *supra* note 9.

69 James Guszcza, Iyad Rahwan, Will Bible, Manuel Cebrian, & Vic Katyal, *Why We Need to Audit Algorithms*, HARV. BUS. REV. (Nov. 28, 2018); Hacohen, *supra* note 1, at 365.

[70] Hacohen, *supra* note 1, at 365.

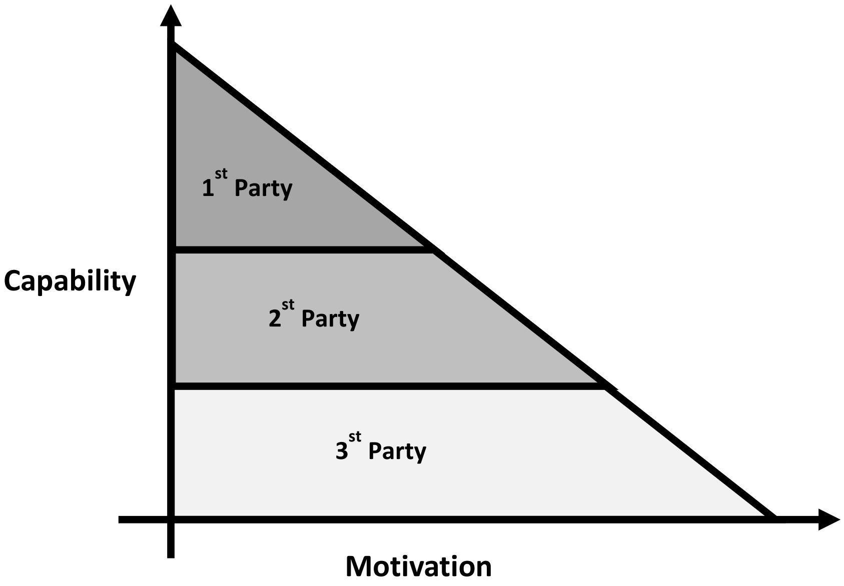

Figure 1: The inverse relationship between technical capability and business motivation to conduct algorithmic auditing.

A. First-Party Audits

First-party audits are audits that platforms conduct of their own products.[71] Platforms perform first-party audits mainly to maintain the workability of their products, promote responsible AI policies, and mitigate legal liability risks.[72] Platforms conduct algorithmic audits with various data analytics and machine learning tools, sometimes even by soliciting input directly from their users.[73]

Because they have uninterrupted access to the raw materials of the algorithmic auditing process—user data and machine learning algorithms—the platforms are the most capable entities to conduct algorithmic audits on an ongoing basis, including in real time.[74] External auditors lack comparable access to the platforms' data and algorithms because the platforms defend these assets zealously through legal and technological means.[75]

For this reason, first-party audits provide policymakers with a rare opportunity to peek into what would otherwise be highly secretive and inaccessible systems.[76] The transparency that first-party audits provide may also trigger public interest and scrutiny, thereby setting the scene for follow-up reviews by external stakeholders.[77]

[71] DRCF, *supra* note 55; *See e.g.,* Miranda Sissons & Ian Levine, *A Closer Look: Meta's First Annual Human Rights Report*, META, July 14, 2022.

[72] *See e.g.,* Kay Firth-Butterfield, Miriam Vogel, *5 ways to avoid artificial intelligence bias with 'responsible AI'*, WORLD ECONOMIC FORUM, (Jul. 5, 2022).

[73] *See, infra* DRCF, section IV.A; Hacohen, Competition *supra* note 1 (define data-driven experimentation).

[74] *Supra* note 69.

[75] Goodman & Trehu, *supra note* 64. For example, platforms avoid open-source code in their systems, embrace restricted licensing terms to prevent data scraping, and zealously prosecute attempts to reverse-engineering their algorithms.

[76] HETAN SHAH, ALGORITHMIC ACCOUNTABILITY. PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A: MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES 376, 2128 (2018),s

[77] *See generally* James Guszcza, Iyad Rahwan, Will Bible, Manuel Cebrian, and Vic Katyal, *Why We Need to Audit Algorithms*, HARV. BUS. REV. (Nov. 28, 2018). External audit can also incite internal auditing in attempt to contradict the finding of the external audit. This adversary is socially desirable because it produce valuable information and trigger policy debate. Consider the COMPAS example. Jon

Nevertheless, the value of first-party audits is naturally limited.[78] Conducted by the audited platforms themselves, first-party audits are prone to be affected by the platforms' own business agendas and interests.[79] Given this conflict of interest, first-party audits are inherently unreliable.[80] Moreover, first-party audits are also restricted to the platform developers' perspectives and limited abilities to predict how their products will be used in practice. For example, while Microsoft's chatbot Tay was carefully examined by the Microsoft team and was trained only on "modeled, cleaned, and filtered" data, its users could still manipulate it to produce racist remarks within hours of its public release.[81]

B. Second-Party Audits

Second-party audits are conducted by external auditors hired by the audited company.[82] Their structure offers a middle ground between first-party and third-party audits. On the one hand, second-party auditors have access to the platforms' data and system designs (similarly to first-party auditors). On the other hand, second-party auditors are still considered at least external and partially independent entities, often bounded by professional or ethical standards (similarly to third-party auditors). Second-party audits are essentially a rough compromise between the platforms' trade secrecy interests and society's transparency and oversight interests, and as such have the virtue of "push[ing] the industry to be more transparent."[83]

However, the social value of second-party audits is limited. Second-party auditors are likely to be responsive to the platforms that hired them, which stains their impartiality. Moreover, the platforms may prevent second-party auditors from sharing proprietary information relevant to the public[84] or

Kleinberg, Sendhil Mullainathan, & Manish Raghavan, *Inherent Trade-Offs in the Fair Determination of Risk Scores,* PROCEEDINGS OF INNOVATIONS IN THEORETICAL COMPUTER SCIENCE (ITCS), 2017 HTTPS://ARXIV.ORG/ABS/1609.05807. *But see* Jacob Metcalf, Ranjit Singh, Emanuel Moss & Elizabeth Anne Watkins, *Witnessing Algorithms at Work: Toward a Typology of Audits*, DATA & SOCIETY: POINTS (Aug. 11), https://points.datasociety.net/witnessing-algorithms-at-work-toward-a-typology-of-audits-efd224678b49 (arguing that audit rebuttal conflicts push social discussion into technical unhelpful debates).

[78] Elettra Bietti, *From Ethics Washing to Ethics Bashing: A Moral Philosophy View on Tech Ethics*, 2(3) JOURNAL OF SOCIAL COMPUTING (Sep. 2021); Inioluwa Deborah Raji, Peggy Xu, Colleen Honigsberg & Daniel Ho, *Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance*, CS.CY. 1 (2022); *Closing the Gap*, *supra* note 63.

[79] Metcalf, Singh, Moss & Watkins, *supra* note 77; *Witnessing Algorithms at Work: Toward a Typology of Audits*, DATA & SOCIETY: POINTS (Aug. 11), https://points.datasociety.net/witnessing-algorithms-at-work-toward-a-typology-of-audits-efd224678b49.

[80] But platforms may still bound themselves to certain codes of ethics. Brent Mittelstadt, for example, surveyed the field in 2019 and found at least 84 AI ethics initiatives publishing frameworks. *see* Brent Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI*, NATURE MACHINE INTELLIGENCE 1, (Nov. 2019). Another fruitful source of objectives is the UN Guiding Principles Reporting Framework, which provides human rights-related goals for businesses, and is the metric that Meta has used to audit its own products. *See* Shift, UN GUIDING PRINCIPLES REPORTING FRAMEWORK, (Jul. 24, 2022). 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. *See* EUROPEAN COMMISSION, ETHICS GUIDELINES FOR TRUSTWORTHY AI, (Jul. 24, 2022).

[81] James Vincent, *Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day*, THE VERGES (Mar. 24, 2016)

[82] *See e.g.* Christo Wilson et. al. *Building and Auditing Fair Algorithms: A Case Study in Candidate Screening*, FACCT '21 (Mar. 1–10, 2021), https://evijit.github.io/docs/pymetrics_audit_FAccT.pdf. The Facebook's civil rights audit—while not explicitly related to algorithms—illustrates the limits of second party auditing. *See* Mark Latonero & Aaina Agarwal, *Human Rights Impact Assessment for AI: Learning from Facebook's Failure in Myanmar*, CARR CENTER DISCUSSION PAPER SERIES, HARVARD UNIVERSITY (2021), https://carrcenter.hks.harvard.edu/publications/human-rights-impact-assessments-ai-learning-facebook%E2%80%99s-failure-myanmar; Facebook's Civil Rights Audit, *Facebook's Civil Rights Audit Report—Final Report*, META July 8, 2020.

[83] *See* Schellmann, *supra* note 88 (citing Pauline Kim, a law professor at Washington University in St. Louis). *But see* Sloane, *supra* note 88 ("This 'collaborative audit' sets a dangerous precedent.")

[84] See e.g., Hilke Schellmann, *Auditors are testing hiring algorithms for bias, but there's no easy fix*, MIT. TECH. REV. (Feb. 11, 2021); https://www.technologyreview.com/2021/02/11/1017955/auditors-testing-ai-hiring-algorithms-bias-big-questions-remain/(noting

from proposing certain meaningful solutions to the harms they identify.[85] For these reasons, second-party audits (along with first-party audits) may provide false assurance that the audited algorithms perform better than they genuinely do.[86] Commentators call this problem "audit washing" (or "ethics washing").[87] For example, in a second-party audit conducted for the job interviewing company Pymetrics, the auditors cleared the company's algorithms of allegations of gender and racial bias. However, they ignored other concerns that may affect the company algorithm's validity.[88] Similarly, HireVue marketed its employment algorithm as having passed a second-party civil rights audit. However, external reviewers later called the auditors' independence and the audit's scope into question.[89]

C. Third-Party Audits

Third-party audits are external audits conducted by independent auditors who have no formal relationship with the audited company. Third-party auditors may be academic researchers, regulators, journalists, or users. Almost by definition, third-party auditors cannot rely on having undisrupted access to the systems they seek to scrutinize.[90] Instead, third-party auditors must rely on indirect verification techniques or trust the platforms' willingness to provide information.

Major technology platforms voluntarily used to provide at least partial access to their data through their platforms' application program interfaces (APIs).[91] Over the years, third-party auditors took advantage of this access and created massive databases, such as the Twitter archive at the Library of Congress.[92] These archives formed a solid basis for algorithmic auditing. Nevertheless, the platforms' embracement of external oversight did not last long.[93] Following the 2018 Cambridge Analytica scandal, Facebook began shutting down hundreds of thousands of applications that used the platforms' API to extract public and personal data.[94] Other companies followed Facebook's lead, and the supply of third-party auditing by academics and civil society groups was heavily restricted.[95]

that, "[t]he auditors were editorially independent but agreed to notify Pymetrics of any negative findings before publication."); Metcalf, Singh, Moss & Watkins, *supra* note 77.

[85] *See e.g.,* Newley Purnell, *Facebook is Stifling Independent Report on its Impact in India, Human Rights Groups Say*, THE WALL STREET JOURNAL, November 12, 2021.

[86] Goodman & Trehu, *supra note* 64.

[87] Julian Jaursch, *Why The EU Needs To Get Audits For Tech Companies Right*, TECHDIRT, (Aug. 19, 2021); Bietti, *supra* note 78; Goodman, *supra* note 6.

[88] Mona Sloane, *The Algorithmic Auditing Trap*, MEDIUM (Mar. 17, 2021) (noting that the second party audit "Interrogated the Pymetrics system for racial and gender bias, but left unexamined the claim that gameplay performance was predictive of job performance."); Susanna Vogel, *What does an "audit" for bias in automated hiring tech really mean?*, HR BREW, (Apr. 1, 2022), https://www.hr-brew.com/stories/2022/04/01/what-does-an-audit-for-bias-in-automated-hiring-tech-really-mean; Schellmann, *supra* note 84. *Cf.* Goodman, *supra* note 6 (noting that Facebook's commissioned human rights impact assessment "focused solely on the United States, at a time when Facebook's human rights record in non-US and non-Anglophone countries was undergoing substantial scrutiny.")

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

[90] Metcalf, Singh, Moss & Watkins, *supra* note 77.

[91] Anat Ben-David, *Counter-archiving Facebook*, 35(3) EUROPEAN JOURNAL OF COMMUNICATION 249 (2020).

[92] Ben-David, *supra* note 91, at 252.

[93] Issie Lapowsky, *Platforms vs. PhDs: How tech giants court and crush the people who study them*, PROTOCOL (Mar. 19, 2021), https://www.protocol.com/nyu-facebook-researchers-scraping.

[94] Ben-David, *supra* note 91, at 250.

[95] *Id.*

With no access to the platforms' APIs, third-party auditors had to develop alternative auditing measures.[96] One such measure is called "sock puppet audits."[97] In sock puppet audits, third-party auditors use real or fake user accounts (sometimes with automated agents or "bots") to scout the platform's ecosystem for information.[98] Researchers have successfully used sock puppet audits in various areas, such as news ranking, [99] video content diversity,[100] and content curation.[101]

Third-party auditing may also be user-based.[102] The rest of this article explores this option.

IV. User-Based Algorithmic Auditing

As producers of the algorithms' input and consumers of the algorithms' output, users are uniquely positioned to detect and reflect algorithmic harm.[103] Nevertheless, users' engagement in the algorithmic auditing process, either to assist other third-party auditors or to make assessments of their own, is surprisingly limited. This section first offers a new typology to describe user-based algorithmic auditing as a special, underdeveloped subcategory of third-party auditing. Next, this section explores the challenges that lead to the limited scale of user-based algorithmic audits and offers policy interventions to make such audits more common.

A. Typology

This section introduces a new typology to explain user-based algorithmic auditing. The proposed typology classifies user-based auditing into two main categories and subcategories. The first distinction is between *user-assisted* and *user-driven* algorithmic auditing approaches. User-assisted algorithmic audits

[96] Ben-David, *supra* note 91, at 255. Inspired by early methods of the social scientific "audit study," Sandvig et al. proposed a taxonomy to summarize different algorithm auditing methods and research designs, Sandvig et. al., *supra* note 11.

[97] Sandvig et. al., *supra* note 11, at 13.

[98] *Id.*

[99] Emma Lurie & Eni Mustafaraj, *Opening up the Black Box: Auditing Google's Top Stories Algorithm*, 32 FLAIRS. 376 (2019).

[100] Aleksandra Urman, Mykola Makhortykh & Roberto Ulloa, *Auditing Source Diversity Bias in Video Search Results Using Virtual Agents*, CS.IR. 1 (2021) (Another example is a video search results audit which examines whether video search outputs are subjected to source diversity bias. The method used to collect the data was a set of software virtual agents simulating user browsing behavior and recording its outputs).

[101] Nathan Bartley, Andres Abeliuk, Emilio Ferrara & Kristina Lerman, *Auditing Algorithmic Bias on Twitter*, WEBSCI '21. 65 (2021): (Additionally in the Twitter recommendation audit the researchers used sock puppet method to quantify the impact of algorithmic curation on the information users see. The bots only observe and collect the data from public accounts (to avoid privacy problems) and it does not create any new data (like, share etc.). The results demonstrate 3 types of bias: popularity and exposure bias); Eduardo Hargreaves, Claudio Agosti, Daniel Menasché, Giovanni Neglia, Alexandre Reiffers-Masson & Eitan Altman, *Biases in the Facebook News Feed: a Case Study on the Italian Elections*, CS.SI. 1 (2018) (researchers used bots and Internet browser extension to study Facebook News Feed algorithm during the Italian elections).

[102] Alicia DeVos, Aditi Dhabalia, Hong Shen, Kenneth Holstein & Motahhare Eslami, *Toward User-Driven Algorithm Auditing: Investigating users' strategies for uncovering harmful algorithmic behavior*, CHI '22. 1 (2022) [DeVos, Towards]; Le, Spina, Scholer & Chia, *supra* note 1; DANAË METAXA, JOON SUNG PARK, RONALD E. ROBERTSON, KARRIE KARAHALIOS, CHRISTO WILSON, JEFF HANCOCK & CHRISTIAN SANDVIG, AUDITING ALGORITHMS: UNDERSTANDING ALGORITHMIC SYSTEMS FROM THE OUTSIDE IN (2021) [hereinafter SYSTEMS FROM THE OUTSIDE IN]; Hong Shen, Alicia Devos, Motahhare Eslami, & Kenneth Holstein, *Everyday Algorithm Auditing: Understanding the Power of Everyday Users in Surfacing Harmful Algorithmic Behaviors*, PROC. ACM HUM. COMPUT. INTERACT. 5, CSCW2, Article 433 (October 2021) [DeVos, Everyday].

[103] User-based auditing may also help auditors avoid legal barriers. For example, The CFAA restrictions on unauthorized use are unlikely to be triggered by a genuine human user interacting with online platforms, when platforms accepts new user accounts from anyone on the Internet. Sandvig et. al., *supra* note 11, at 14; Christian Sandvig, Kevin Hamilton, Karrie Karahalios & Cedric Langbort, *Auditing algorithms: Research methods for detecting discrimination on internet platforms*, DATA AND DISCRIMINATION: CONVERTING CRITICAL CONCERNS INTO PRODUCTIVE INQUIRY, 22, 4349-4357 (2014).

are third-party audits that rely on users' input in their auditing analysis. The other category, user-driven algorithmic audits, are audits that the users themselves initiate and govern.

A second classification breaks down algorithmic audits into *supervised* and *unsupervised* categories. Drawing on traditional machine learning terminology,[104] supervised audits are designed to address specific predefined inquiries and research hypotheses, and unsupervised audits are frameworks for scrutinizing algorithmic systems with no guiding hypotheses.[105] The following subsections explore these categories in greater detail.

a. *User-Assisted Algorithmic Auditing*

User-Assisted algorithmic auditing refers to third-party audits that rely on users' data as input for their auditing analysis. These audits can be supervised or unsupervised, depending on the auditors' goals. The following subsections explore these two options.

i. Supervised

In supervised user-assisted audits, third-party auditors solicit user input to address specific research hypotheses.[106] For example, academic researchers or civil society groups may turn to the platforms' users to collect the data they need at the requisite scale to conduct effective algorithmic audits.[107] Auditors solicit users' assistance in various ways. The most intuitive approach is through a surveying format.[108] For instance, in a study conducted at Syracuse University, researchers interviewed TikTok users identified with the LGBTQ community about their experiences using TikTok's For You page.[109] Building on these users' input, the researchers gained valuable insights concerning TikTok's content curation algorithm.[110]

The survey-based method is helpful mainly to sense the users' sentiments towards the algorithms' performance; however, it is limited in its analytic value. Survey-based studies suffer from serious validity issues because they lack the benefits of manipulation or random assignment to conditions that conventional experimental studies utilize.[111] Thus, auditors using this approach cannot rigorously infer causality from any given results.[112]

[104] *See .e.g.*, Devin Soni, *Supervised vs. Unsupervised Learning*, MEDIUM (Mar. 22, 2020), https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d (last visited Jul 20, 2023).

[105] Aline Iramina, Maayan Perel, & Niva Elkin-Koren, *Paving the Way for the Right to Research Platform Data* (June 19, 2023). Available at SSRN: https://ssrn.com/abstract=4484052 or http://dx.doi.org/10.2139/ssrn.4484052 (invoking the need to facilitate exploratory research).

[106] The term "users" is defined broadly in this definition. Many times auditors will refer to "everyday" users in their supervised user-assisted audits but not every time. For example in an auditing study of political ad targeting criterial in online platforms, Matias et al. used "testers" to place ads. These "testers" played the role of advertisers which are also considered platform users given that these platforms operate in multi-sided markets. Matias, *supra* note 64. Users can aggregate data to solve social problems unrelated to algorithmic auditing. These endeavors are sometimes called "data cooperatives" or "social machines." *See generally* NIGEL SHADBOLT ET. AL, THE THEORY AND PRACTICE OF SOCIAL MACHINES (Springer, 2019).

[107] Sandvig et. al., *supra* note 11

[108] *Id.* at 11.

[109] Ellen Simpson & Bryan Semaan. *For You, or For "You"?: Everyday LGBTQ+ Encounters with TikTok. Proc.* 252 ACM HUM.-COMPUT. INTERACT. 4, CSCW3, (December 2020).

[110] *Id.*

[111] Sandvig et. al., *supra* note 11, at 11.

[112] *Id.*

A more accepted user-assisted auditing method is crowdsourcing.[113] In supervised user-assisted crowdsourcing audits, researchers hire users to work on specific tasks they assign them in advance. Researchers have used crowdsourcing techniques to investigate numerous algorithmic issues, including search engine personalization,[114] price steering and discrimination in e-commerce sites,[115] and content diversity.[116] The crowdsourcing methodology allows auditors to collect user data at a large scale, which enables in-depth studies of even marginal irregularities.[117]

In recent years, the rise of popular crowdsourcing platforms such as Amazon Mechanical Turk (AMT) has made the crowdsourcing methodology accessible to researchers in numerous disciplines, even to those with little training in empirical studies.[118] Unsurprisingly, Christian Sandvig and

[113] *See e.g.* Jakub Mikians, László Gyarmati, Vijay Erramilli & Nikolaos Laoutaris, *Crowd-assisted search for price discrimination in e-commerce: First results*, PROCEEDINGS OF THE NINTH ACM CONFERENCE ON EMERGING NETWORKING EXPERIMENTS AND TECHNOLOGIES. 1 (2013).

[114] Binh Le, Damiano Spina, Falk Scholer & Hui Chia, *A Crowdsourcing Methodology to Measure Algorithmic Bias in Black-Box Systems: A Case Study with COVID-Related Searches*, in ADVANCES IN BIAS AND FAIRNESS IN INFORMATION RETRIEVAL 43 (Ludovico Boratto, Stefano Faralli, Mirko Marras & Giovanni Stilo eds., 2022) (A study of ADM+S and (ARC) regarding misinformation spread in search queries linked to COVID-19. The study originated in RMIT University and The University of Melbourne, and used the Amazon Mechanical Turk crowdsourcing platform. Fifty users volunteered and were asked to accept the task, fill a pre-task question about their age/gender/education, manually run a provided query using Google search, save the SERP as HTML and upload the file to the MTurk. The results showed that the composition of search results differ in different countries and between different crowd workers who used the same queries. Also, different search results occur depending on whether the user used a positive or a negative query formulation.); Ronald E. Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer & Christo Wilson, *Auditing Partisan Audience Bias within Google Search,* 148 ACM Vol. 2, 1 (2018) (Research done in Northeastern University in 2018, to study partisan audience bias within Google search. The researchers recruited demographically diverse 187 participants from the crowdsourcing platforms "Prolific" and "Crowdflower." The participants installed a custom browser extension to test different queries. Then the researchers compared the results given to them by the users to a unique dataset containing the sharing propensities of registered Democrats and Republicans on Twitter. The results showed little evidence for the "filter bubble" hypothesis (meaning that personalization increases the partisan bias of web search) and that results at the bottom of Google SERPs were more left leaning. Yet this study limitation is that most of the users were white males, and their political preferences were not balanced, and that could cause a deviation from expectation).

[115] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, & Christo Wilson, *Measuring price discrimination and steering on e-commerce web sites,* In PROCEEDINGS OF THE 2014 CONFERENCE ON INTERNET MEASUREMENT CONFERENCE. 305–318 (2014) (A study from Northeastern University in 2014, investigating 16 of the top e-commerce sites for instances of price steering and discrimination in retail, hotels and cars. For that purpose, the researchers used web scraping techniques and hired 100 Amazon MTurk users (from the US only) through ("HIT"). Which means the researchers used both Crowdsourcing platform and "Sock puppet" auditing methods. The users answered a brief survey of previous accounts in the e-commerce sites and then configure their web browser to use a Proxy Auto-Config (PAC) file provided by the researchers. That allowed the researchers receive the data the users are seeing and compare the results to the fake accounts (with different characteristics) for checks and data controlling. The results points to 7-8 sites that implement personalization.); Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, & Christo Wilson, *Measuring price discrimination and steering on e-commerce web sites,* In PROCEEDINGS OF THE 2014 CONFERENCE ON INTERNET MEASUREMENT CONFERENCE (2014) 305–318 (similar).

[116] Tim Glaesener, *Exploring Siri's Content Diversity Using a Crowdsourced Audit*, 128 JOURNAL OF DIGITAL SOCIAL RESEARCH Vol. 4 No.1 (2022) (Researchers recruited 134 US-based users from Amazon Mechanical Turk. User were asked to report Siri's answers to five questions. Although the results of this audit were analyzed, the researcher pointed out that they are very limited due to the small number of participants and lack of ability to prove that Siri's answers reported were accurate.); Jack Bandy & Nicholas Diakopoulos, *Auditing News Curation Systems: a Case Study Examining Algorithmic and Editorial Logic in Apple News,* RESEARCHGATE (Aug. 1st, 2019), https://www.researchgate.net/publication/334866841_Auditing_News_Curation_Systems_A_Case_Study_Examining_Algorithmic_and_Editorial_Logic_in_Apple_News (user auditing was needed because the Apple News app not only lacked public API's but it also implemented strong security measures to prevent data scraping. This research used AMT workers who received a Human Intelligence Task ("HIT") to collect screenshots of Apple News.)

[117] Alex Abdo, Ramya Krishnan, Stephanie Krent, Evan Welber Falcón & Andrew Keane Woods, *A Safe Harbor for Platform Research*, KNIGHT INSTITUTE AT COLUMBIA UNIVERSITY (Jan 19, 2022), https://knightcolumbia.org/content/a-safe-harbor-for-platform-research.

[118] Daniel B. Shank, *Using Crowdsourcing Websites for Sociological Research: The Case of Amazon Mechanical Turk*, AM SOC. 47 (2016). The fact that crowdsourced/collaborative auditing rely on crowd workers, who do not necessarily represent the demographics of a given algorithmic system's user base. That is especially the case when the researchers are using AMT, since most AMT users come from the United States and India. That might cause major blind-spots in the research. Also, Amazon is itself a major platform that is suseptible to various algorimic (and other) abuses, so relying on such plafrom as a research tool is somewhat problematic.

colleagues consider the crowdsourcing method "the most useful and promising for future work in this area."[119]

In theory, first-party auditors (the platforms themselves) can also conduct supervised user-assisted auditing.[120] Deborah Raji and colleagues, for example, proposed a framework for self-auditing, which leverages past user data to identify failures in algorithmic performance.[121] In this vein, the platforms, like third-party auditors, may use survey and crowdsourcing methods as part of their self-auditing. Meta, for example, applies a crowdsourcing approach to allow users to report inappropriate social media content and then uses this input to update its content curation algorithms.[122] Twitter recently used a survey approach when it solicited user input before launching its new deepfake filtration algorithm.[123]

Platforms may even compensate users for providing their input for algorithmic auditing. This approach builds on the well-established practice of "bug bounties," whereby platforms pay hackers to detect vulnerabilities and security failures in their released software.[124] For instance, Attenberg and Provost explored a game-like methodology that invites users to bring forth instances they believe will cause algorithmic systems to fail.[125] Users who "beat the machine" by flagging unexpected flaws in the audited systems are compensated for their success.

Nevertheless, this article does not consider the platforms' utilization of user data in its definition of user-based algorithmic auditing.[126] As explained in Part II.A., first-party auditing is biased by definition and cannot be trusted.[127] For this reason, what might be considered "user-assisted platform

[119] Sandvig et. al., *supra* note 11, at 15.

[120] This article does not consider this option seriously because of the limited value of first-party audits. *See* Part III. To the extent that the algorithmic auditing process is designed to remedy the quasi-feudal power balance between platforms and their users, first-party auditing is inherently suspected. *See* EXPLORING LEGAL MECHANISMS FOR DATA STEWARDSHIP, ADA LOVELACE INSTITUTE 33 (Mar. 2021), https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/ ("The imbalances of power or ability of individuals and groups to act in ways that define their own future create a data environment that is in some ways akin to the feudal system which fostered the development of trust law."). *See generally* Ben-David, *supra* note 91 (invoking the same sentiment).

[121] Raji et, *supra* note 63.

[122] *What happens when I report something to Facebook? Does the person I report get notified?*, HELP CENTER, FACEBOOK, https://www.facebook.com/help/103796063044734. More generally, Cabrera et al. showed that platforms might effectively recognize and audit systematic failures in their algorithmic systems by way of crowdsourcing user failure reports. Angel Alexander Cabrera, Abraham J. Druck, Jason I. Hong, Adam Perer, Discovering and Validating AI Errors With Crowdsourced Failure Repoets, 5 PROC. ACM HUM.-COMPUT. INTERACT. 5, CSCW2, 425:1 (October 2021).

[123] Mariel Soto Reyes, *Twitter is soliciting user feedback to build out its deepfake policy*, INSIDER (Nov. 13, 2019) https://www.businessinsider.com/twitter-solicits-user-input-to-shape-deepfake-policy-2019-11. Facebook also tried a model for crowdsourcing for verifying fake news but it is a very limited experiment. *See* Bernhard Clemm, *Analysis | Facebook wants its users to drive out fake news. Here's the problem with that.*, WASHINGTON POST, February 1, 2018, https://www.washingtonpost.com/news/monkey-cage/wp/2018/02/01/facebook-wants-to-drive-out-fake-news-by-having-users-rate-news-outlets-credibility-heres-the-problem-with-that/ (last visited Jun 25, 2018); Amber Jamieson & Olivia Solon, *Facebook to begin flagging fake news in response to mounting criticism*, THE GUARDIAN (Dec. 15, 2016), https://www.theguardian.com/technology/2016/dec/15/facebook-flag-fakenews-fact-check.

[124] Thomas Maillart, Mingyi Zhao, Jens Grossklags, and John Chuang, *Given enough eyeballs, all bugs are shallow? Revisiting Eric Raymond with bug bounty programs*. 3(2) JOURNAL OF CYBERSECURITY 81–90 (2017).

[125] Joshua Attenberg, Panos Ipeirotis, & Foster Provost, *Beat the Machine: Challenging Humans to Find a Predictive Model's "Unknown Unknowns*, 6(1) J. DATA AND INFORMATION QUALITY (mar 2015). This work belongs to a broader body of work has explored the design of crowd pipelines, interactive visualizations, and interfaces to support crowd workers in searching for and making sense of algorithmic errors. See e.g., Ángel Alexander Cabrera, Abraham Druck, Jason I. Hong, & Adam Perer, *Discovering and Validating AI Errors With Crowdsourced Failure Reports*. 5 CSCW2 (2021); Jina Suh, Soroush Ghorashi, Gonzalo Ramos, Nan-Chen Chen, Steven Drucker, Johan Verwey, & Patrice Simard, *AnchorViz: Facilitating Semantic Data Exploration and Concept Discovery for Interactive Machine Learning*. 7 ACM TRANS. INTERACT. INTELL. SYST. 10, 1, (Aug. 2019).

[126] *See supra* note 120.

[127] *See supra* Part II.A.

self-auditing" is more accurately classified as self-serving experimentations to improve algorithmic performance than a genuine investigation of algorithmic social harms.[128]

ii. Unsupervised

In unsupervised user-assisted audits, auditors crowdsource data to explore possible algorithmic harms without specific guiding hypotheses in mind. To conduct unsupervised algorithmic audits, auditors must have access to the algorithm input and output data.[129] Platforms have unrestricted access to both data sources and their algorithms, which allows them to conduct unsupervised algorithmic audits at will. As noted above, this privilege is usually unavailable to third-party auditors, because platforms do not usually share enough of their datasets with external stakeholders. For instance, services such as Google Trends[130] and Facebook Ad Archive[131] provide third-party auditors with only partial access to their algorithms' input data (search queries and advertisers' ads, respectively) but with no comparable access to the algorithms' output data (search results and targeted ad information, respectively).[132]

Third parties may try to engage in unsupervised algorithmic audits by collaborating with the platforms' users. As producers of the algorithms' input and consumers of the algorithms' output, users are invaluably positioned to reconstruct datasets that platforms usually keep secretive. Thus, by collaborating with users to crowdsource the relevant data at scale, third-party stakeholders can create large enough datasets to form a solid basis for unsupervised algorithmic audits.[133] Anat Ben-David helpfully labeled this endeavor "counter-archiving."[134]

Many researchers crowdsource user data to conduct unsupervised audits. AlgorithmWatch, Ad Observatory, and Who Targets Me are all crowdsourced counter-archives that support unsupervised audits of Instagram and Facebook's ad targeting algorithms.[135] Researchers made these archives by

[128] Indeed, calling these self-serving algorithmic optimization processes "auditing" is probably stretching this term beyond its reasonable boundaries. *See generally,* Hacohen, Competition, *supra* note 1 (explain that platforms exploit users' data for positive as well as negative ends).

[129] Ayelet Gordon-Tapiero, Katrina Ligett & Alexandra Wood, *The Case for Establishing a Collective Perspective to Address the Harms of Platform Personalization*, 42 VAN. J. OF ENT. & TECH. (2022). *See generally* Jenna Burrell, How the machine 'thinks': Understanding opacity in machine learning algorithms, BIG DATA & SOCIETY, (Jun.1, 2016).

[130] Simon Rogers, *What Is Google Trends Data—And What Does It Mean*, GOOGLE NEWS LAW (July 1, 2016), https://medium.com/google-news-lab/what-is-google-trends-data-andwhat-does-it-mean-b48f07342ee8.

[131] Ad Library, META, https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=IL&media_type=all.

[132] Hacohen, Competition, *supra* note 1. The non-profit imitative Who Targets Me, and various counter-archives for Facebook ads are trying to fill out this gap.

[133] In addition, such reconstruction is has the benefit of being more authentic than archives that are based on input form the platforms. Critics of API-based research have pointed at the risks of making truth claims using social media data without taking into account that instead of mediating social or political phenomena, these data are originally created to meet specific corporate goals and ideologies. Ben-David, *supra* note 91, at 255; John NA and Nissenbaum A, *An agnotological analysis of APIs: Or, disconnectivity and the ideological limits of our knowledge of social media*, 35(1) THE INFORMATION SOCIETY 1–12 (2019); Marres N and Gerlitz C, *Interface methods: Renegotiating relations between digital social research, STS and sociology*, 64(1) THE SOCIOLOGICAL REVIEW 21–46 (2016).

[134] Ben-David, *supra* note 91, at 256 ("It differs from other methods of data mining and public data sharing or dumping, in the sense that it does not simply make data sets available, but rather consciously incorporates appraisal, facilitates use and is transparent about its definition of 'publicness' and provenance." [internal citations omitted]).

[135] Ad Observer originated at New York University, and studied the spread of disinformation on Facebook, focusing on advertisements. NYU Researchers relied on the data delivered to them by Facebook as part of the transparency tools assigned to researchers, and on data collected by consenting Facebook users. The users installed a browser extension which enabled them to voluntarily submit their own ad targeting data as they browsed Facebook. The NYU researchers could access everything the user could see from his browser. Similarly, AlgorithmWatch's goal was to monitor, with the help of 1,500 volunteers, Instagram's news feed algorithm to achieve a better

offering users browser extensions that automatically scraped their social media data. Then the researchers aggregated that data in searchable databases that served as the foundation for unsupervised audits by third-party auditors, such as academics and journalists.[136]

Meta eventually blocked all these projects for allegedly violating its platform's terms of service and other applicable regulations.[137] To deal with these restrictions, some researchers created counter-archives with tools that evade the platforms' anti-scraping policies. Anat Ben-David, for example, created the *Meturgatim* platform to enable unsupervised audits of Facebook's political ad targeting practices in Israel. To overcome Facebook's preventive measures, Ben-David built the *Meturgatim* archive by crowdsourcing screenshots of political ads coming from anonymous users and then converting the information from these images into searchable textual data.[138]

Government entities also engage in user-assisted unsupervised auditing.[139] For example, the Australian Research Council (ARC) and Centre of Excellence for Automated Decision-Making and Society (ADM+S) crowdsourced data from over 1,000 Google users for the sake of an unsupervised audit of the Google search engine algorithm.[140]

b. *User-Driven Algorithmic Auditing*

User-driven algorithmic auditing refers to the processes that users themselves initiate and govern. Similarly to user-assisted auditing methods, user-driven methods can be supervised or unsupervised. The present subsection investigates both options.

understanding of how the algorithm prioritizes pictures and videos. The results show that the algorithm preferred politicians or political content and representations of body-image. The research was supported by the European Data Journalism Network and by the Dutch foundation SIDN. *See* Nicolas Kayser-Bril, *AlgorithmWatch forced to shut down Instagram monitoring project after threats from Facebook*, ALGORITHM WATCH (Aug. 13, 2021), https://algorithmwatch.org/en/instagram-research-shut-down-by-facebook/#:~:text=Newsletters,AlgorithmWatch%20forced%20to%20shut%20down%20Instagram%20monitoring%20project%20after%20threats,to%20hold%20them%20to%20account; Le, Spina, Scholer & Chia, *supra* note 114; Abdo, Krishnan, Krent, Falcón & Woods, *supra* note 117.

[136] *See e.g.,* Nancy Watzman, *How journalists use NYU Ad Observatory to report on digital political advertising in elections*, MEDIUM (Sep. 6, 2022), https://medium.com/cybersecurity-for-democracy/how-journalists-use-nyu-ad-observatory-to-report-on-digital-political-advertising-in-elections-2a296dcc0adf.

[137] Abdo, Krishnan, Krent, Falcón & Woods, *supra* note 117; Lapowsky, *supra* note 93; Kayser-Bril, *supra* note 135.

[138] Ben-David, *supra* note 91, at 259.

[139] *Cf.* Priscilla M. Regan, *A Design for Public Trustee and Privacy Protection Regulation*, 44 SETON HALL LEGIS. J. 487, 506 (2020); Jennifer Shkabatur, *The Global Commons of Data*, 22 STAN. TECH. L. REV. 354, 393 (2019) (arguing that existing regulatory agencies can assume this role); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 25 (2014) (suggesting that expert technologists from the FTC or FCC could be granted access to private scoring algorithms "to test them for bias, arbitrariness, and unfair mischaracterizations."). Freeman Engstrom et al., *supra* note **Error! Bookmark not defined.** at 88–90 (discussing the pros and cons of regulatory collaboration between the government and private entities). See generally Colleen Honigsberg et al., *Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance*, ARTIFICIAL INTELLIGENCE, ETHICS & SOC'Y (2022) https://arxiv.org/pdf/2206.04737.pdf (favoring third party auditing by the government or organization to internal auditing).

[140] Le, Spina, Scholer & Chia, *supra* note 114; Australian Research Council (ARC) and Centre of Excellence for Automated Decision-Making and Society (ADM+S) aimed to understand the different "Search Engine Results Page" (SERPs) in different ages in Australia. This project started in 2021 and since then collected more than 350 million search results from 1000+ citizens who downloaded a plug-in to their browser. The project gains national support because of its broad search engines auditing like Chrome, Edge, and Firefox browsers on 48 topics. The project is based upon the non-profit organization AlgorithmWatch from Germany (2017). *Automated Decision-Making and Society* [hereinafter ADM+S], https://www.admscentre.org.au/news/.

i. Supervised

Supervised user-driven audits refer to audits that users initiate in response to some specific hypothesis or "folk theory" about harmful algorithmic behavior.[141] In their day-to-day encounters with algorithms, users may encounter puzzling algorithmic behaviors that would lead them to initiate an audit.[142] Users would then collaborate to share and analyze the relevant data together. Sophie Bishop labeled these attempts to share information among users "algorithmic gossip."[143] Ultimately, users might attempt to publish their findings to raise public awareness and seek to trigger follow-up scrutiny by other third parties.

For example, one user suspected that Apple's credit scoring system was gender-biased, after finding that his wife's credit score differed from his. He then tweeted his suspicion on social media, which led other users to join the audit and gather evidence. This grassroots investigation gradually expanded until the federal government finally picked it up.[144] In another case, a woman suspected gender biases in the Google Translate algorithm after she tried translating gender-neutral pronouns from foreign languages to English and got gender-biased results (such as a male engineer and a female nurse).[145] Similarly, a group of LGBTQ+ YouTubers discovered that YouTube's commercial algorithm demonetizes queer-related content after noticing that their advertising earnings are not proportional to their audience traffic.[146]

Some user-driven algorithmic audits are very successful. Most notably, a user-driven audit of Twitter's image-cropping algorithm successfully uncovered systematic bias against Black users and managed to attract significant public backlash.[147] Twitter notified that it tested the service before launching but failed to detect this bias.[148] The audit was so successful that Twitter had no choice but to follow up with another audit. This time, the company even offered the users a bounty for detecting

[141] See Karizat et al., *Algorithmic Folk Theories and Identity: How TikTok Users*, PROC. ACM HUM.-COMPUT. INTERACT., Vol. 5, No. CSCW2, Article 305 (Oct. 2021).

[142] DeVos, Everyday, *supra* note 102, at 433:24 (noting that, "Through their day-to-day interactions with an algorithmic system, everyday users are particularly well positioned to detect these types of behaviors that emerge in real-world contexts of use, in the presence of complex social dynamics, and in the changing norms and practices of using algorithmic systems over time."). past literature suggests, many harmful machine behaviors are challenging to detect outside of situated contexts of use. *See* Henriette Cramer, Jean Garcia-Gathright, Aaron Springer, and Sravana Reddy, *Assessing and addressing algorithmic bias in practice*, INTERACTIONS 25, 6 58–63 (2018); Batya Friedman and Helen Nissenbaum, *Bias in computer systems*. ACM TRANSACTIONS ON INFORMATION SYSTEMS (TOIS) 14, 3, 330–347(1996); Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudik, and Hanna Wallach, *Improving fairness in machine learning systems: What do industry practitioners need?*. In PROCEEDINGS OF THE 2019 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS. 1–16; Michael A Madaio, Luke Stark, Jennifer Wortman Vaughan, and Hanna Wallach, *Co-designing checklists to understand organizational challenges and opportunities around fairness in AI.* In PROCEEDINGS OF THE 2020 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS. 1–14 (2020); Nick Seaver, *Algorithms as culture: Some tactics for the ethnography of algorithmic systems*, BIG DATA & SOCIETY 4, 2 (2017). https://doi.org/doi:10.1177/2053951717738104.

[143] Sophie Bishop, *Managing visibility on YouTube through algorithmic gossip*, 21 (11-12) NEW MEDIA AND SOCIETY 2589 (2019).

[144] Neil Vigdor, *Apple Card Investigated after Gender Discriminations Complaint,* THE NEW YORK TIMES (Nov. 10th, 2019), https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html.

[145] Parmy Olson, *The Algorithm That Helped Google Translate Become Sexist*, FORBES (Feb. 15th, 2018), **https://www.forbes.com/sites/parmyolson/2018/02/15/the-algorithm-that-helped-google-translate-become-sexist/?sh=48116527daa2**.

[146] Aja Romano, *A group of YouTubers is trying to prove the site systematically demonetizes queer content,* VOX (Oct. 19th, 2019) **https://www.vox.com/culture/2019/10/10/20893258/youtube-lgbtq-censorship-demonetization-nerd-city-algorithm-report.**

[147] DeVos, Everyday *supra* note 102.

[148] DeVos, Everyday *supra* note 102

additional flaws in its system.[149] In another user-driven audit, users revealed Yelp's manipulation of its review filtering algorithm, possibly to force businesses to pay more for getting better ratings. The federal government also picked up on this investigation, which resulted in nearly 700 FTC reports.[150]

In a comprehensive study, Alicia DeVos and colleagues investigated how users develop their auditing hypothesis by inviting users to brainstorm and uncover hidden biases in Google Images' search engine algorithm.[151] The users in the study found over 150 actual biases, highlighting the value of user-driven audits as a methodological approach.[152] The study also triggered the opening of discussion forums and data visualization platforms such ImageNet Roulette[153] and Search Atlas[154] to boost user-driven algorithmic scrutiny.[155]

Besides raising public awareness and additional scrutiny from other stakeholders,[156] user-driven audits can also inspire algorithmic resistance or activism.[157] Users may follow up on their auditing efforts with "data strikes"[158] or engage in "data poisoning"[159] or "data leveraging"[160] campaigns, all in

[149] This crowdsourcing call was naturally for programmers, some of them may be Twitter users, but not necessarily. Rumman Chowdhury & Jutta Williams, *Introducing Twitter's first algorithmic bias bounty challenge*, TWITTER BLOG (July 30th. 2021), https://blog.twitter.com/engineering/en_us/topics/insights/2021/algorithmic-bias-bounty-challenge.

[150] Motahhare Eslami, Kristen Vaccaro, Min Kyung Lee, Amit Elazari Bar On, Eric Gilbert & Karrie Karahalios, *User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms*, CHI '19 494 (2019).

[151] DeVos, Towards *supra* note 102.

[152] *Id.*

[153] ImageNet Roulette was a simple online interface developed by artists and researchers to support users in exploring and interrogating the input/output space of an image captioning model trained on the ImageNet dataset. This project provoked discussions on social media, as users shared findings and hypotheses, and sometimes built upon each other's observations. *See* Kate Crawford and Trevor Paglen, *Excavating AI: The politics of images in machine learning training sets*. AI & SOCIETY (2021), 1–12.

[154] Search Atlas, which allows users to explore and easily compare Google search results as if they were located in different countries. *See* Rodrigo Ochigame and Katherine Ye., *Search Atlas: Visualizing Divergent Search Results Across Geopolitical Borders.* (2021), 1970–1983. https://doi.org/10.1145/3461778.3462032

[155] DeVos, Everyday, *supra* note 102,

[156] "Algorithmic awareness" is a vital component in the development of algorithmic folk theories by users, but further knowledge of algorithms and their effects on users' social media experiences may also result in "algorithm disillusionment" if users' algorithmic expectations do not match the actual processes occurring. *See* Motahhare Eslami, Karrie Karahalios, Christian Sandvig, Kristen Vaccaro, Aimee Rickman, Kevin Hamilton, & Alex Kirlik, *First I "like" it, then I hide it: Folk Theories of Social Feeds*, IN PROCEEDINGS OF THE 2016 CHI CONF. ON HUMAN FACTORS IN COMPUTING SYSTEMS. ACM, SAN JOSE CALIFORNIA USA, 2371–2382. https://doi.org/10.1145/2858036.2858494; Motahhare Eslami, Sneha R. Krishna Kumaran, Christian Sandvig, and Karrie Karahalios, *Communicating Algorithmic Process in Online Behavioral Advertising* In PROCEEDINGS OF THE 2018 CHI CONF. ON HUMAN FACTORS IN COMPUTING SYSTEMS - CHI '18. ACM PRESS, MONTREAL QC, CANADA, 1–13. https://doi.org/10.1145/3173574.3174006.

[157] *See* DeVos, Everyday, *supra* note 102. For example, algorithmic folk theories can influence users' actions through online hashtag campaigns to expose and resist possible platform changes. Michael Ann Devito, Darren Gergle, & Jeremy Birnholtz, *Algorithms ruin everything: #RIPTwitter, Folk Theories, and Resistance to Algorithmic Change in Social Media.* In PROCEEDINGS OF THE 2017 CHI CONF. ON HUMAN FACTORS IN COMPUTING SYSTEMS. ACM, DENVER COLORADO USA, 3163–3174. https://doi.org/10.1145/3025453.3025659. *See generally,* Nancy Ettlinger, *Algorithmic affordances for productive resistance*, BIG DATA & SOCIETY 1 (Jan. 2018); Quentin Grossetti, Cédric du Mouza, & Nicolas Travers, *Community-Based Recommendations on Twitter: Avoiding the Filter Bubble.* IN WEB INFORMATION SYSTEMS ENGINEERING – WISE 2019.

[158] Nicholas Vincent, Brent Hecht, & Shilad Sen, *"Data Strikes": Evaluating the Effectiveness of New Forms of Collective Action Against Technology Platforms*, In PROCEEDINGS OF THE WEB CONFERENCE 2019.

[159] Marco Barreno, Blaine Nelson, Russell Sears, Anthony D Joseph, & J Doug Tygar, C*an machine learning be secure?*. IN PROCEEDINGS OF THE 2006 ACM SYMPOSIUM ON INFORMATION, COMPUTER AND COMMUNICATIONS SECURITY. 16–25; Battista Biggio, Blaine Nelson, & Pavel Laskov, *Poisoning attacks against support vector machines*. ARXIV PREPRINT ARXIV:1206.6389 (2012).

[160] Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, & Brent Hecht, *Data leverage: A framework for empowering the public in its relationship with technology companies*. IN PROCEEDINGS OF THE 2021 ACM CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY. 215–227.

an attempt to affect algorithmic behaviors or to pressure platforms to perform changes to their algorithms.[161]

ii. Unsupervised

Unsupervised audits are unguided explorations of user-crowdsourced databases in search of unidentified algorithmic harms. The only difference between the user-driven unsupervised audits discussed here and the user-assisted unsupervised audits described in subsection IV.A.a.ii. is that the former are initiated and governed by the users themselves (bottom-up), while the latter is initiated and governed by third-party auditors (top-down).

That said, unsupervised user-driven algorithmic auditing is nearly impossible in today's environment. Because data is a networked good, users wishing to conduct unsupervised audits face a severe collective action problem.[162] To reach the scale needed, users must pool their data together and build large enough datasets to adequately support unsupervised auditing. Cooperation of such a scale is challenging to execute, especially amongst users spread globally, across international borders.[163]

B. Analysis: Policy Interventions to Promote User-Based Algorithmic Auditing

Users interact with algorithms daily and are deeply affected by algorithmic actions. As such, users are uniquely able to detect and flag algorithmic harm.[164] Moreover, because users know which data they provide to the algorithms as input (for example, what Facebook knows about them[165]) and which data they receive from the algorithms as output (for example, which ads Facebook shows them[166]), as a collective, they can effectively reconstruct the databases that underline the algorithms' actions. Nevertheless, users' engagement in the algorithmic auditing process is quite limited. The following subsections explore the main challenges of user-based auditing and introduce preliminary thoughts on policy interventions.

a. *User-Assisted Algorithmic Auditing*

[161] Motahhare Eslami, Kristen Vaccaro, Karrie Karahalios & Kevin Hamilton, *Be Careful Things Can Be Worse Than They Appear: Understanding Biased Algorithms and Users' Behavior around Them in Rating Platforms,* PROCEEDINGS OF THE ELEVENTH INTERNATIONAL AAAI CONFERENCE ON WEB AND SOCIAL MEDIA. 62 (2017).

[162] *See generally* Hacohen, competition, *supra* note 1; Exploring legal mechanisms for data stewardship, *supra* note 120 at 62.

[163] *See generally* Hacohen, policy, *supra* note 1.

[164] In researcher driven audits, one of the main issues arises in the early stages, while researchers struggle to detect and mitigate harmful biases in the algorithms due to cultural blind-spots, and unanticipated circumstances in the contexts of where and when a system is used. In addition, some biases will only occur in the presence of real-world social or cultural dynamics. Thus, auditing the code only is not sufficient enough. DeVos, Everyday, *supra* note 102 at 433:23 ("everyday algorithm auditing leverages the lived experiences of everyday users."); Meg Young, Lassana Magassa, and Batya Friedman, *Toward inclusive tech policy design: A method for underrepresented voices to strengthen tech policy documents,* 21(2) ETHICS AND INFORMATION TECHNOLOGY, 89–103 (2019) ("suggests that existing approaches often fail to involve auditors with relevant cultural backgrounds and lived experience that are critical to detect sensitive harms")

[165] Andrew Hutchinson, *Facebook Looks to Provide More Data Transparency with Revamped 'Access Your Information' Tool*, SOCIAL MEDIA TODAY (Jan. 12, 2021) https://www.socialmediatoday.com/news/facebook-looks-to-provide-more-data-transparency-with-revamped-access-your/593246/

[166] Help Center, *Why am I seeing ads from an advertiser on Facebook*?, META. https://www.facebook.com/help/794535777607370.

i. Supervised

Academic researchers, civil society groups, and investigative journalists who conduct supervised user-assisted audits are opening themselves to a wide array of legal risks. For instance, audits that involve collecting user data or scraping data directly from the platforms' web pages can violate the platforms' Terms of Service (TOS) and expose auditors to liability.[167] Facebook's terms of service are illustrative. Facebook prohibits users from accessing or collecting data from its platform with automated means without the platform's prior permission.[168] Facebook also prohibits users from creating or using accounts solely for research purposes.[169] The platform even prohibits people from facilitating or supporting *others* in doing these things.[170] Lastly, Facebook purports to prohibit *former* users from engaging in any of these conducts.[171] Facebook can change its terms to make data collection lawful for particular purposes; but even then, the platform may change its course on a whim.[172]

Other legal hurdles include the liability that third-party auditors may face for security breaches or intellectual property and privacy violations.[173] The Computer Fraud and Abuse Act (CFAA), for example, imposes liability on whoever "accesses a protected computer without authorization"[174] or "exceeds authorized access."[175] These sections threaten third-party auditors because courts did not clearly define what constitutes "unauthorized access" online.[176] Auditors also risk liability under the European GDPR if the data they collect implicate third-party users or publishers who have not formally consented to the use of their information.[177] These risks may chill third parties from conducting algorithmic audits or, at the minimum, force them to adopt auditing methods that are less risky, but more labor intensive and less efficient.[178]

[167] *See generally* Alex Abdo et. al., *A Safe Harbor for Platform Research*, The Knight Institute (Jan. 19, 2022), https://knightcolumbia.org/content/a-safe-harbor-for-platform-research#:~:text=Importantly%2C%20the%20safe%20harbor%20protects,platforms%20from%20erecting%20new%20ones.

[168] Facebook, Terms Of Service, Facebook § 3(2)(1) (Last visited or Aug. 6, 2023), https://www.facebook.com/legal/terms?paipv=0&eav=AfYrIkHMJ_Gwdm1WljpkLv-xPx88pgtBykcP58jH8ICmIZomdygN8OlC3CUclrxqauc&_rdr

[169] *Id.*

[170] *Id.* at § 3(2).

[171] *Id.* at § 4(2).

[172] *See* Abdo, *supra* note 167.

[173] *See* Abdo, *supra* note 167.

[174] Section (a)(2)(C).

[175] Section (a)(4).

[176] Annie Lee, *Algorithmic Auditing and Competition under the CFAA: The Revocation Paradigm of Interpreting Access and Authorization*, 33 Berkeley Technology Law Journal 1307 (2018). However, a new ruling by the US Supreme Court in a recent lawsuit, Van Buren v. United States, resulted with the statement "research aimed at uncovering whether online algorithms result in racial, gender, or other discrimination does not violate the Computer Fraud and Abuse Act". United States v. Detroit Timber & Lumber Co., 200 U. S. 321, 337. In another case, Sandvig v. Barr, a federal judge ruled that journalists and researchers may access data for research studies, without violating the terms-of-service. Goodman & Trehu, *supra note* 64.

[177] Abdo, Krishnan, Krent, Falcón & Woods, *supra* note 117. However, the General Data Protection Regulation has an exemption for the processing of personal data "for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes," Regulation 2016/679 of the European Parliament and of the Council of Apr. 27, 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), art. 89, 2016 O.J. (L 119) 17 (EU).

[178] Goodman & Trehu, *supra note* 64 (noting that "the risk of legal liability will shape how the audit is conducted"); Ben-David, *supra* note 91 (Compared with other methods of data extraction, the rather 'primitive' screenshot offers limited analytical possibilities and is difficult to collect.").

To aid supervised user-assisted algorithmic auditing, governments should create a safe harbor for researchers conducting user-assisted algorithmic audits.[179] Safe harbor legislation would immunize user-assisted third-party auditors from legal liability, eliminating the deterrent caused by the companies' terms of service, the CFAA, and state-law analogs to the CFAA.[180] Recently, the Knight Institute at Columbia University proposed a legislative safe harbor for platform research, which was incorporated nearly verbatim into the draft of the Platform Accountability and Transparency Act.[181]

Policymakers should also oblige platforms to actively share information with researchers as mandated by the Platform Accountability and Transparency Act and the EU's Digital Services Act.[182] However, data-sharing obligations alone will not be sufficient, because the scope of these mandates will necessarily be limited[183] and because third-party auditors should be able to scrutinize and verify the platforms' compliance with these obligations.[184] Therefore, policymakers should implements transparency and safe harbor regulations in tandem.

ii. Unsupervised

Other than the legal risks involved in platforms' research, third-party auditors may also be dissuaded from pursuing unsupervised user-assisted audits simply because of the high costs, efforts, and ongoing maintenance these large-scale projects require. A significant cost factor in the execution of such projects is attributed to the fact that users' data is a networked good.[185] As such, users are not highly incentivized to participate in these projects unless a critical mass of users have already joined in.[186]

Governments can overcome these challenges by subsidizing unsupervised user-assisted algorithmic auditing.[187] Alternatively, governments could even undertake the role of conducting user-assisted algorithmic auditing themselves.[188] To that end, governments could appoint third-party auditing entities and bestow them with legal fiduciary responsibility toward the platforms' users.[189] Then, governments could encourage (even oblige[190]) users to share data with these entities. Designated APIs may facilitate such data-sharing responsibilities.

[179] *See* Abdo, *supra* note 167; Goodman & Trehu, *supra note* 64 (More protections for adversarial audits carried out by researchers or journalists without a company's consent may be required.").

[180] *See* Abdo, *supra* note 167;

[181] *See* Abdo, *supra* note 167; Platform Accountability and Transparency Act, *supra* note 41.

[182] *See supra* notes 41, 48.

[183] Iramina, Perel, & Elkin-Koren, *supra* note 105.

[184] *See* Abdo, *supra* note 167 (noting that "Researchers must be able to ensure that datasets are comprehensive, and relying on the platforms' own disclosures is not enough.")

[185] *See generally*, Hacohen Competition, *supra* note 1 (explaining that user data is a networked good); RadicalxChange, *infra* note 214; Exploring legal mechanisms for data stewardship, *supra* note 120 at 106 (referencing RadicalxChange for the notion of market failure in aggregating user data).

[186] Hacohen*, id;* RadicalxChange, *id.*

[187] *Cf.* REGULATING AI IN THE UK: STRENGTHENING THE UK'S PROPOSALS FOR THE BENEFIT OF PEOPLE AND SOCIETY, ADA LOVELACE INSTITUTE 38 (Jul. 2023) ("Recommendation 10: Create formal, funded channels to involve civil society organizations, particularly those representing vulnerable groups. . . ").

[188] *See* Hacohen, Policy, *supra* 1, at 365.

[189] *See* Hacohen, Policy, *supra* 1, at 365.

[190] At a minimum for user data that only implicates the privacy interests of commercial third parties (such as advertisers).

Driven by similar considerations, Aziz Huq has suggested that governments should create "public trusts" for their citizens' data.[191] As Huq explains, "[a]n asset in public trust is owed and managed by the state . . . .The state can permit its use, and even allow limited alienation, provided that doing so benefits a broad public rather than a handful of firms."[192] Huq's suggestion takes inspiration from some initiatives that have already support trusted intermediaries to govern user data.[193] For instance, the Spanish city of Barcelona requires all companies within its borders to share data with the city's local platform, Decidem, where user-generated data uses "will be subject to public debate and decision."[194] Similarly, the Silicon Valley Data Trust integrates information streams from benefits agencies, child protection bureaus, schools, and education technology companies to create a "well-managed regional data trust."[195]

Government-appointed third-party auditing entities would serve a similar function as the data trust model suggested by Huq, but for the more limited purposes of algorithmic auditing and research. This proposal can draw inspiration from the various regulatory approaches, such as the German Network Enforcement Act (NetzDG),[196] that delegate auditing responsibilities to intermediary entities.[197] However, like most existing global regulations, the NetzDG empowers auditors to gather the relevant data from the audited platforms, not directly from the platforms' users.[198]

b. *User-Driven Algorithmic Auditing*

iii. Supervised

Supervised user-driven algorithmic auditing is currently suffering from severe problems of coordination and expertise.[199] Most app users are not statisticians or computer scientists; thus, they lack the technical skills to execute algorithmic audits. Unsurprisingly, when Twitter solicited user input as part of its "bias bounty" program,[200] it aimed its call at participants with the relevant technical expertise, as platforms usually do in traditional cybersecurity bounty settings.[201]

In addition, because data is a networked resource, identifying data-driven algorithmic harms requires a very broad analytic perspective that individuals or small communities lack.[202] Consider gig

[191] *Cf.* Aziz Z. Huq, *The Public Trust in Data*, 110 WASH. L. J. 333, 335 (2021)

[192] *Id.* At 333.

[193] *Id.* at 337–39.

[194] *Id.* at 337 (referencing Amy Lewin, Barcelona's Robin Hood of Data, SIFTED (Nov. 16, 2018), https://sifted.eu/articles/barcelonas-robin-hood-of-data-francesca-bria/ [https://perma.cc/7QMF-ZNR4]).

[195] *Id.* at 337 (referencing SILICON VALLEY REGIONAL DATA TRUST, https://www.svrdt.org [https://perma.cc/5QLE-US8D].

[196] *See, e.g.,* Ben Wagner et al., *Regulating Transparency? Facebook, Twitter and the German Network Enforcement Act*, BARCELONA, SPAIN: ACM CONFERENCE ON FAIRNESS ACCOUNTABILITY AND TRANSPARENCY (FAT*) (Jan, 2020).

[197] *Id.*

[198] *Id.* Other examples include the French Digital Act, which grants the government statistical authorities the right to access platform-held information under certain conditions, the European DSA which compels platforms to share information about how their algorithms function with academia and civil society groups, the US Platform Transparency and Accountability Act which authorizes the Federal Trade Commission to compel data platforms to share some of their data with independent researchers, and many more. See Hacohen, policy, *supra* note 1. at 371.

[199] DeVos, Everyday, *supra* note 102

[200] *See* Chowdhury & Williams, *supra* note 149.

[201] DeVos, Everyday, *supra* note 102 at 433:22.

[202] DeVos, Everyday, *supra* note 102 at 433:21 (noting that, "in order to establish that a harmful bias truly exists and is worth addressing, it is necessary to compare multiple instances and demonstrate the existence of statistical patterns."); DeVos, Towards, *supra* note

workers, who cannot tell how algorithms manage their work unless they resort to a wage comparison app that helps them collectively reveal systematic inequalities.[203] This problem grows exponentially when considered on the global scale in which the platform algorithms operate.[204] For example, Namibian users cannot know whether Facebook targets them with dark patterns more invasively than, say, Italian users, without being able to access and analyze the Italian users' data alongside their own (collective) data.[205] Even if the platforms' users are perceptive enough to suspect algorithmic harm from their limited perspective, they still need access to a broader perspective to corroborate their concerns. For example, after finding that a web search for her name: "Latanya Sweeney" yielded arrest record ads, Sweeney suspected that the Google Search algorithm may systematically associate Black-sounding names with a criminal record.[206] To corroborate her concern, it was necessary to search for many Black-sounding names and compare statistically the results against those for non-Black-sounding names.[207]

To address these challenges, governments could create designated digital forums for users to share their concerns and hypotheses of algorithmic harms publically, and to coordinate their auditing efforts with other users. Besides helping users to flag and coordinate their collaborative auditing efforts, these platforms could also incorporate the services of expert auditors who would provide information and guidance to users.[208]

The government could also facilitate reputational rewards based on a community-driven ranking of marginal effort and contribution.[209] This approach can be modeled based on other communities of collaborative production, such as Wikipedia. Governments could even go further to support users'

102("User-driven audits thus far have occurred in situations where users notice potential issues within algorithmic systems, whether by stumbling upon or by actively searching for problematic algorithmic behaviors. This limits the relevance of user-driven auditing to algorithms that are directly visible to users to interact with, neglecting many algorithmic systems that are offline and invisible to users.").

[203] Aarian Marshall., *Gig Workers Gather Their Own Data to Check the Algorithm's Math*, WIRE https://www.wired.com/story/gig-workers-gather-datacheck-algorithm-math/.

[204] *Cf.* COMMISSION COMMUNICATION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS, A EUROPEAN STRATEGY FOR DATA, 6, COM (2020) 66 final (Feb. 19, 2020) ("Fragmentation between Member States is a major risk for the vision of a common European data space and for the further development of a genuine single market for data."); Hacohen, Competition, *supra* note 1 (manuscript at 27) (arguing that in the digital platform environment, economies of scale support concentration on a global rather than national or regional scale).

[205] Gordon-Tapiero et al., *supra* note129, at 9 (explaining that only platforms have the ability to fully perceive the "picture of [] personalization landscape").

[206] Latanya Sweeney, *Discrimination in online ad delivery*, QUEUE 11, 3, 10–29 (2013).

[207] *Id.*

[208] DeVos, Everyday, *supra* note 102 at 433:21 ("To help steer everyday auditors' ongoing efforts in more productive directions, designers might invite relevant experts to continuously monitor community discussions and intervene as needed." For example, at any stage of an everyday audit, a product team might weigh in and provide feedback, based on technical knowledge of their products, regarding the plausibility of user-generated hypotheses.').

[209] DeVos, Everyday, *supra* note 102 at 433:20 ("we might imagine tailoring or augmenting discussion channels with mechanisms that are explicitly designed to support everyday auditing. For example, a discussion platform designed for everyday auditing might allow community members to upvote specific algorithmic behaviors that other community members have reported, in order to collectively surface the most severe reports, or reports that may benefit from further discussion."); DeVos, Everyday, *supra* note 102 at 433:22 ("we could imagine organizing everyday auditors along a number of major roles, such as (1) Initiators, who focusing on identifying problematic algorithmic behaviors; (2) Generalizers, who examinine instances flagged by initiators, and use these to inform hypotheses about the scope of the issue and possible underlying causes; (3) Amplifiers, who help
share and broadcast what has been discovered to others, raising awareness of findings, hypotheses, and relevant discussions so far; and (4) Synthesizers, who transform the outputs of this iterative auditing process into summaries for others to build on."); DeVos, Towards, *supra* note 102 at user driven ("The creation of new platforms like these apps, with affordances that support collective investigation and sensemaking, can provide avenues for the recognition of harmful algorithmic behaviors that may otherwise remain invisible.")

participation by providing monetary awards based on their active involvement in auditing processes.[210] Such incentive schemes could be modeled after the various whistleblower regimes that already exist at the federal level, such as the False Claims Act or the Securities and Exchange Commission bounty program.[211] Similarly, platform users could flag suspected algorithmic harms to a regulator, provide a logical basis and some initial evidence for their hypothesis, and get a comparable bounty in cases where the regulator pursues with an investigation.

Notably, even in the absence of government support, supervised user-driven auditing might also be supported by market developments and technology. For example, users can already use machine learning procedures to scale up their hypotheses of algorithmic harm.[212] Indeed, Michelle Lam and colleagues introduced an end-user auditing tool called IndieLabel*,* which allows users to project their insights from a small number of users onto a much larger test set.[213] Additional technological solutions in the near future might ameliorate the collective action problems associated with supervised user-driven algorithmic audits, however the problems of lack motivations and technical expertise are likely to subsist.

iv. Unsupervised

The same cooperation challenges discussed also apply to unsupervised user-driven algorithmic audits, but much more intensively. To get valuable insights from unsupervised audits, users must create massive counter-archives of data and keep maintaining them, ideally in real time. These efforts will likely be significant, given the lack of standardized frameworks to support such cooperation; legal barriers that only augment friction (such as privacy); and, most notably, the lack of financial or reputational incentives for users to cooperate. These challenges intensify with the size and multi-nationality of audited platforms.

The government could foster unsupervised user-driven algorithmic auditing by helping users to coordinate.[214] In this vein, scholars and legislators have proposed creating trusted intermediaries that will allow users to facilitate data pooling.[215] For example, Sylvie Delacroix and Neil D. Lawrence offer a model of "bottom-up data trusts," which would upset the current top-down data governance

[210] DeVos, Everyday, *supra* note 102 at 433:22 ("another major design opportunity is to explore mechanisms that incentivize everyday auditors to have more active and sustained participation.")

[211] Phillip & Cohen, *SEC whistleblower program: Overview, rewards and protections*, PHILLIP & COHEN https://www.phillipsandcohen.com/sec-and-cftc-whistleblower-info/#:~:text=An%20SEC%20whistleblower%20(Securities%20and,and%20protection%20from%20job%20retaliation.

[212] Michelle S. Lam, Mitchell L. Gordon, Danaë Metaxa, Jeffrey T. Hancock, James A. Landay, & Michael S. Bernstein, *End-User Audits: A System Empowering Communities to Lead Large-Scale Investigations of Harmful Algorithmic Behavior* | PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION, https://dl.acm.org/doi/10.1145/3555625 (last visited Jul 20, 2023).

[213] Michelle S. Lam, Mitchell L. Gordon, Danaë Metaxa, Jeffrey T. Hancock, James A. Landay, & Michael S. Bernstein, *End-User Audits: A System Empowering Communities to Lead Large-Scale Investigations of Harmful Algorithmic Behavior* | PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION, https://dl.acm.org/doi/10.1145/3555625 (last visited Jul 20, 2023).

[214] On the technical fronts projects such as the Ocean Protocol, Streamr, and Swash are pushing into this direction. *See* Leon Erichsen, Matt Prewitt, *Solving The Social Dilemma: The Data Freedom Act,* RADICALXCHANGE https://www.radicalxchange.org/media/blog/solving-the-social-dilemma/; *see also* Hacohen, policy, *supra* note 1, at 365.

[215] *See* Hacohen, policy, *supra* note 1, at 365.

structure and enhance users' collective bargaining power.[216] Initiatives such as the Decode project[217] and RadicalxChange[218] pursue a similar vision. Recent legislation in the European Union also pushes in this direction.[219] The Data Governance Act creates an intermediary layer of trusted "data sharing services," which would handle sharing users' data without utilizing it for other purposes.[220] Similar proposals are also being considered in the UK.[221]

While these efforts aim to democratize data governance structures and loosen the concentration in digital markets, they would also provide a framework for creating counter-archives and facilitating unsupervised algorithmic audits. Auditing is an unrecognized prerequisite for any user-centered data governance regime vision.[222] Without auditing, users cannot know the ills of an existing data ecosystem and cannot collectively bargain to change them.

Thus, the government can draw on this emerging "data trusts" literature to facilitate the creation of bottom-up "auditing trusts."[223] Governments can define the qualifying conditions for these intermediaries and even employ a vetting supervision process similar to the one already created in section 40 of the Digital Service Act.[224] Governments can even encourage (or oblige) platforms to support these intermediary entities with APIs and to facilitate and streamline users' participation with behavioral nudges such as opt-in defaults.[225]

Table 2: Challenges to user-based algorithmic auditing and suggestions to help overcome them

	User-Assisted	**User-Driven**
Supervised	• Challenges: Legal risks for third-party auditors (contract, intellectual property, privacy). • Suggestions: Safe harbor for auditors.	• Challenges: Lack of motivation and expertise. • Suggestions: Financial incentives (awards or subsidies) and guidance.
Unsupervised	• Challenges: Collective action.	• Challenges: Collective action.

[216] Sylvie Delacroix & Neil D. Lawrence, Bottom-Up Data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance, 9 INT'L DATA PRIVACY L. 236, 238 (2019); *See also* PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL ON EUROPEAN DATA GOVERNANCE (Data Governance Act), COM (2020) 767 final (Nov. 25, 2020), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767 [https://perma.cc/3KD9-Z8W5] [hereinafter DGA];

[217] The Decode Project, DECODE, https://decodeproject.eu/what-decode.html.

[218] Erichsen & Prewitt, *supra* note 214.

[219] *See generally* Hacohen, policy, *supra* note 1.

[220] DGA, *supra* note 216.

[221] *See Exploring legal mechanisms for data stewardship, supra note 185.*

[222] Indeed, this point is underemphasized in the data trust literature. For an overview see Hacohen, policy *supra* note 1

[223] Cf. Delacroix & Neil D. Lawrence, *supra* note 216.

[224] For analysis *see e.g.,* John Albert*, A guide to the EU's new rules for researcher access to platform data, Algorithm Watch*, https://algorithmwatch.org/en/dsa-data-access-explained/.

[225] *See generally* RICHARD H. THALER, CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (Penguin, 2009)

	- Suggestions: Financial incentives (awards or subsidies); Institutional means (facilitating top-down auditing trusts).	- Suggestions: Technological means (AI), Institutional means (facilitating bottom-up auditing trusts).

V. Conclusion

The rise of digital platforms driven by AI and cloud computing has brought significant social, economic, and political power to companies like Meta, Google, and Amazon. While these platforms offer valuable services to users, they also pose risks of algorithmic bias, discrimination, and manipulation. Policymakers are actively seeking ways to hold these platforms accountable, and algorithmic auditing has emerged as a key tool in this endeavor.

Existing regulatory frameworks have leaned towards first-party auditing, where platforms self-assess their algorithms. However, this approach has its flaws, as commercial interests may not align with societal well-being. To address this conflict of interests, the regulatory focus is now shifting towards third-party auditing, with independent entities overseeing the platforms. However, even this approach has limitations, as the information provided for auditing originates and is controlled by the platforms themselves.

In response to these challenges, this article proposes a unique and underutilized approach: user-based algorithmic auditing. This approach empowers the platforms' users to initiate or govern the auditing process, or to collaborate with independent auditors. By involving users directly, the dependency on platforms is minimized, enabling a more impartial assessment of algorithmic systems. Moreover, user-based audits can corroborate the information provided by platforms, providing a more balanced, holistic understanding of potential harms.

This article explored the current regulatory frameworks for algorithmic auditing, analyzed existing auditing approaches, and introduced user-based algorithmic auditing. The article concluded with a few suggestions that would empower user-based algorithmic to improve platform-driven research and governance. As contemporary society increasingly depends on digital platforms, attaining reliable audits of their potential harms becomes a pressing priority.