June 12, 2023

Response to the National Telecommunications and Information Administration's Request for Comment on AI Accountability Policy

Data & Society Research Institute (Data & Society or D&S) is pleased to submit a response to the Request for Comment published by the National Telecommunications and Information Administration (NTIA) on Artificial Intelligence (AI) system accountability measures and policies.

Our organization is an independent, nonprofit research institute studying the social implications of data-centric technologies and automation. We are working to produce empirical research that challenges the power asymmetries created and amplified by technology in society.

Mandating public-facing documentation, impact assessments, and means to redress are essential to create an accountable AI ecosystem. The public cannot be expected to protect themselves from systems of which they have no statutory right to inspect or demand changes. The government must ensure these rights, as well as abide by and uphold the principles of the Blueprint for an AI Bill of Rights to ensure that Americans are protected from algorithmic harm and discrimination

Our comment addresses the purposes, possibilities, and limitations of algorithmic accountability mechanisms and the structures and regulations needed to realize a true democratic system of algorithmic accountability in the United States.[1] We hope this aids NTIA's aim to "develop a productive AI accountability ecosystem" and to create a report on AI accountability development.

Q1. What is the purpose of AI accountability mechanisms such as certifications, audits, and assessments?

We believe it is important to make distinctions between the types of AI accountability documentation practices, such as audits, assessments, and certifications. Defining these documentation practices illustrates how certain practices accomplish different goals for the public interest. We propose the following:

> *Audit*: the study of the functioning of a system within the parameters of the system. An audit asks: does the system function appropriately according to a claim made by the developer, according to an independent standard (such as one set by the IEEE, ISO, or NIST), according to terms set in a contract, or according to ethical or scientific terms established by a researcher or a field of researchers?
>
> *Assessments (a/k/a impact assessment)*: the study of the consequences of a system outside of the parameters of the system. An assessment asks: what does the system do to the people, communities, or environment in which it operates? This inquiry should be measured against some pre-established methods or norms such as human rights, ethical development practices, environmental sustainability, social and historical justice, distributive and allocative justice, and community expectations of fairness.
>
> *Certification*: the process of an independent body stating that a system has successfully met some pre-established criteria. Certification asks: has the process that was used to develop and/or deploy this system met the criteria that this body has publicly defined?

Each of these practices has a role to play in an assurance ecosystem, but none can fulfill every possible role. For example, certifications can play the role of making procurement and contracting much smoother. An RFP can specify that a buyer will only purchase an AI system that has received a certain safety or fairness certification from an independent body, and vendors will know in advance what design specifications they must meet. However, that alone cannot satisfy the need of a community to understand and contest whether a proposed system is desirable or just — that purpose is far more appropriate for an impact assessment. Similarly, audits will always be a largely technical exercise, whose utility will exist inside of the engineering and product lifecycle. While the information contained in audits will likely be one aspect of any impact assessment or certification practices, they do not present the additional types of information needed to make democratic decisions.

While AI accountability mechanisms alone cannot solve systemic risks of harm, they can, when done well, be a critical element in identifying and resolving harms. One of the principal aims of accountability measures is "to get the people who build systems to think methodically about the details and potential impacts of a complex project before its implementation, and therefore head off risks before they become too costly to correct."[2] Crucially, regulations for accountability measures that require harm identification provide the opportunity to produce meaningful changes to the internal practices of organizations.

The central task of such accountability mechanisms is to determine what type of algorithmic transparency regimes would provide the public with the requisite knowledge to contest the harms they identify and contend with on a daily basis. Efforts to engage and empower the public around algorithmic systems have drawn on regulation and case law for environmental harms,[3] public nuisance,[4] and regulation of food, drugs, and cosmetics[5] as suitable precedents. What these disparate regimes share in common is that they all are grappling with uncertainty around establishing causes and consequences of harms.

To contend with this uncertainty, more robust forms of accountability measures would require developers to document the expected impacts of such systems, and submit that documentation itself, or a report summarizing the assessment, to a government agency. An ecosystem of organizations providing assurance of AI systems could potentially be operationalized to mediate this accountability relationship between the developer and the regulatory agency. This agency would:

(1) mandate significant public consultation with stakeholders who might be affected by the system; (2) require developers to address harmful impacts that could be ameliorated by changes to the design or deployment of the system; and (3) make aspects of documentation produced by developers publicly available. As proposed, such regulation satisfies a need for greater understanding of how algorithmic systems produce harmful impacts. Crucially, by locating responsibility for overseeing such accountability measures within a federal agency, regulatory approaches have the opportunity — if drafted appropriately — to create conditions for public(s) such as workers and marginalized communities to unite around algorithmic harms they contend with and open possibilities for them to contest whether these accountability measures were followed appropriately in courts.

The AI Bill of Rights provides the opportunity to enable communities contending with algorithmic harms to assert their rights and claim due process. However, the rights formulated under these guidelines can only be actionable within an algorithmic accountability regime that assures that systems have been thoroughly tested for known possibilities of harm prior to their deployment and documentation on their performance on these tests is publicly available.

Q7. Are there ways in which accountability mechanisms are unlikely to further, and might even frustrate, the development of trustworthy AI? Are there accountability mechanisms that unduly impact AI innovation and the competitiveness of U.S. developers?

Accountability mechanisms, if done poorly and if done merely to manage technical risks and vulnerabilities of algorithmic systems, may frustrate the development and deployment of trustworthy AI because they will not account for the full gamut of algorithmic harms.

Algorithmic systems are sociotechnical systems that require assessment methods that can address their simultaneously technical and social dimensions. Overly technical assessments with no accounting for human experience have limited utility. Furthermore, algorithmic impacts arise from algorithmic systems’ actual or potential effects on the world. Assessment methods that do not engage with the world — e.g., checklists or audits based on closed-ended questionnaires for developers — do not foster engagement with real-world effects or the assessment of novel harms. **Put simply, harm must be assessed from the ground up. If people do not have the means to self-identify and articulate their own experiences with AI harms, we anticipate the industry capture of accountability mechanisms.** Even formally independent assessors/auditors can become dependent on a favorable reputation with industry or industry-friendly regulators that could soften their overall evaluations. Conflicts of interest for assessors/auditors should be anticipated and mitigated by alternate funding for assurance work, pooling of resources, or other novel mechanisms for ensuring their independence.

Q9. What AI accountability mechanisms are currently being used? Are the accountability frameworks of certain sectors, industries, or market participants especially mature as compared to others? Which industry, civil society, or governmental accountability instruments, guidelines, or policies are most appropriate for implementation and operationalization at scale in the United States? Who are the people currently doing AI accountability work?

There are examples of AI accountability mechanisms that are currently under discussion, such as use of algorithmic impact assessments in Canada as a self-assessment tool for developers[6] and the proposed EU AI regulations[7] that define risk tiers, prohibit certain applications deemed contrary to human rights, and require conformance audits for higher risk systems prior to deployment. However, **it is important to note that regulatory structures that focus only on the relationship between a regulatory agency and developers for reporting and accepting/approving accountability measures are at high risk of the regulator becoming fully dependent upon the developer for defining and measuring impacts.**

It is inevitable that developers will assess their own systems. The question at hand is whether they alone are permitted to choose the metrics by which their systems are assessed or if the public can exert pressure on the thoroughness and adequacy of the assurance assessments. The design of and regulations around any assurance ecosystem for building trust in AI must take this question into account.

Q16. and Q17. [Paraphrased] How should AI accountability be done properly, by lifecycle and scope (voluntary/mandatory)?

The simplest way to begin building towards AI accountability is to start with documentation on the technical characteristics of a defined model and relevant data. Algorithmic systems are predominantly developed by private companies and are often hidden behind intellectual property and trade secrets protections. These protections contribute to their opacity, especially around how models are developed in the first place and what training data was used. Making technical characteristics of models and data public provides an initial foothold for external, third-party actors as well as the public writ large who wish to pursue accountability by calling attention to the workings of any company's product.

However, developers' concerns around technical documentation and anticipated harms may not align with the harms that concern the public interest. A sole focus on technical documentation would result in over reliance of accountability measures on what developers reveal about their systems, or what can be gleaned from their outputs. This raises two distinct problems: (1) selective disclosure through documentation that leaves the public with little recourse; and (2) the availability of technical documentation on features that may only be of interest to developers and do not provide enough information to understand the relationship between algorithmic impact/harms and system design.

Addressing these problems requires a broader sociotechnical evaluation of the system in which the model is embedded. Such evaluation can happen only after the model is developed; it will begin with taking its technical documentation as the starting point to document how the model will perform in the context of its use within a particular system. In terms of lifecycle, integrating the model within a system happens after the model is developed and tested in some way for its target outcome; this integration work is crucial to a broader sociotechnical evaluation of an algorithmic system. Such evaluations likely will focus only on developer-defined use cases for the system. These developer definitions of use cases are as important as accounts of system performance, because they open the possibility of thinking through the system's potential impact for diverse user populations and personas. While developers may define the use cases of their system narrowly, resulting in inadequate evaluations, the public disclosure of such evaluations can create the grounds for contestation over whether the system was tested thoroughly prior to its deployment.

As illustrated above, there are different **catalyzing events** that can trigger the processes of organizing accountability measures. It is possible to establish clear guidelines around the nature of these events and what they entail. For example, the National Environmental Policy Act mandates an Environmental Impact Assessment when proposed developments receive federal (or certain state-level) funding, or when such developments cross state lines. Similarly, **events such as internal evaluation of a model's performance and integration of a model into an existing or emerging algorithmic system designed for specific use cases can be used by regulators as catalyzing events that trigger requirements of documentation and audits/impact assessments.**

However, it is crucial to specify catalyzing events as moments that can bring about meaningful change in the design of a model and the system that builds on it. The timing of the catalyzing event must account for how and when a system can be altered. For example, in the context of Privacy Impact Assessments, a catalyzing event is prescribed as any point before a system is launched, which leads critics to worry that their results will come too late in the design process to effect change.

In terms of timing, most impact assessments, on one hand, occur *ex ante* before a proposed project is undertaken and/or system is deployed, although they can often also involve ongoing review of how actual consequences compare to expected impacts. Human Rights Impact Assessments, on the other hand, are conducted *ex post*, as a forensic investigation to detect, remedy, or ameliorate human rights impacts caused by corporate activities. Both approaches have merit because impact assessments are not written in stone, and the potential impacts they anticipate (when conducted in the early phases of system deployment) may not be the same as the impacts that can be identified during later phases of system deployment. Additionally, assessments that speak to the scope and severity of impacts may prove to be over- or under-estimated once an algorithmic system is deployed. Similarly, audits can be conducted internally (first-party audits) at any time during the development of a system, *ex ante* by an auditor (second-party audits) before a system is deployed, and *ex post* by an independent actor or researcher (third-party audits) after the system is deployed. **Each of these practices of evaluating the performance of an algorithmic system has its own merits and constraints. To produce meaningful accountability — one in which the public has the ability to contest algorithmic harms in the public domain — they should all work together.**

Public access is crucial for the success of an accountability regime grounded in audits/impact assessments. **Algorithmic accountability policies must specify the level of public access for documentation on audits/impact assessments, which determines who has access to the impact statement reports, supporting evidence, and procedural elements.**

Broad access to impact statements improves an impact assessment's potential to enact changes in system design, deployment, and operation. For environmental impact assessments, public disclosure of an environmental impact statement is mandated legislatively, coinciding with a mandatory period of public comment. For financial impact assessments, fiscal impact reports are usually filed with the municipality as matters of public record, although local regulations vary. Without a strong commitment to make the assessment accessible to the public at the outset, the company may withhold assessments that cast it in a negative light. If too many results are withheld, the public cannot meaningfully protect their interests.

AI accountability measures in the United States would inevitably be industry sector-specific, given that different sectors of the industry such as healthcare, finance, and education are regulated differently. **Rather than trying to define risk of the technology and/or deployment context, policymakers should first look at existing regulations within the sector in which a given algorithmic system is deployed and how they may need to be updated to create conditions for an algorithmic accountability regime.** Existing regulations must be the foundation for new policies around the use of algorithmic systems in a particular sector.

Q25. Is the lack of a general federal data protection or privacy law a barrier to effective AI accountability?

Without comprehensive federal data privacy protections, Americans have few protections from abusive data practices and do not have visibility and agency over how their data is used. Enacting federal data privacy protection will help build trust in AI, particularly algorithmic decision-making models trained on personal information. Comprehensive federal privacy protections should include protections against unfettered data reuse, regulation of the data broker market driving the "inference economy," and heightened scrutiny for sensitive domains, particularly those intersecting with protected classes.

Q30. What role should government policy have, if any, in the AI accountability ecosystem?

Government policy is necessary in order to create an AI accountability ecosystem that truly protects the public. From the rushed deployment of large language models that threaten children's wellbeing,[8] to racist and sexist facial recognition systems that further punitive policing practices,[9] to Americans losing access to critical public benefits due to biased and error-prone automated decision systems,[10] AI systems are inflicting serious harm. But these harms will not resolve themselves without government intervention. First, private technology companies are accountable, first and foremost, to their shareholders. Second, AI threatens usual paths to justice. Automation bias and trade secret protections render it difficult for a victim to prove detrimental harm by an AI system in court.[11] And the opacity of automated systems often forecloses a victim's ability to understand in the first place how an algorithm led to a stop by the police, the denial of public benefits, or online harassment.

We applaud efforts by the White House, such as President Biden's Executive Order on Advancing Racial Equity and Support for Underserved Communities through the Federal Government, to acknowledge the government's role in systemic discrimination and demonstrate its commitment to remedying ongoing harms. But, as companies are increasingly deploying AI systems in ways that impact people's economic opportunities, financial wellbeing, access to housing, and quality of life, the federal government must proactively shape the algorithmic accountability ecosystem.

We encourage NTIA to advance federal AI policy, including needed federal data privacy and algorithmic accountability legislation, regulatory interventions, updated agency practices, and AI-related government research and development efforts to uplift the five principles presented in the White House Office of Science and Technology Policy's Blueprint for an AI Bill of Rights.[12] Without a government AI approach that abides by these principles, we risk creating a mirage of accountability that still leaves Americans victim to algorithmic harm and discrimination.

Our call for active government involvement in the AI accountability ecosystem is not to ignore the voluntary and private industry efforts to further AI accountability. Private technology companies have generated numerous frameworks that can serve as models for documentation requirements, such as datasheets for Datasets[13], Model Cards[14], and disparate impact reporting. Similarly, standards organizations and federal agencies have begun promulgating playbooks for accountable data systems, such as NIST's "AI Risk Management Framework." But, without binding frameworks from federal regulators, these efforts remain voluntary, scattered, and wholly unsynchronized.[15]

Even if leading technology companies wish to conduct assessments of their systems, they are often stymied by the lack of a coherent regulatory vision and reliable market conditions that incentivize industry-wide adoption. It is important that the government enacts an AI accountability regulatory approach that truly assesses the harms of algorithmic systems on historically marginalized communities (instead of only requiring purely technical assessments), ensures civil rights and civil liberties protections against algorithmic discrimination, and provides swift and accessible paths to due process and redress.

Q31. What specific activities should the government fund to advance a strong AI accountability ecosystem?

Both algorithmic impact assessments and algorithmic system audits are key tools that are starting to be used by industry and policymakers to create the conditions for algorithmic accountability. However, as we noted in our response to question 9, the public (in particular historically marginalized communities) must be active participants in the development of assessment metrics and methods of redress. Therefore, there is an opportunity for research and standard setting agencies like NIST and NSF to support research in three critical areas:

- The methodological approaches to assessing the impact of algorithmic decision-making systems, including how these assessments should be conducted in a participatory manner, and how they should remain the same or differ across contexts to achieve the strongest accountability protections;
- The social and economic valuation of audits and impact assessments in terms of the role they play in increasing accountability and mitigating harms; and
- Other approaches to creating the conditions for accountable algorithmic systems that either complement audits and assessments or explore different mechanisms to structure the role of impacted populations in particular and the public writ large in algorithmic accountability.

These three critical areas all require sociotechnical research methods and experts to be given access to AI systems and funding. Sociotechnical research seeks to make often-invisible human, material, and cultural infrastructures visible to better assess the use of technologies in new arenas. Such an approach considers not simply how to best use a technology, but fundamentally *whether a given technology is appropriate in the first place*, and where it fits alongside existing processes of accomplishing work.

Technical research absent broader engagement with experts on society, politics, economy, and culture is likely to reproduce patterns of incomplete, biased, and discriminatory solutions. The US government should lead in funding sociotechnical research that makes the US a leader in technical and qualitative bias and discrimination correction and mitigation methods.

Thank you for the opportunity to offer recommendations to create an accountable AI ecosystem. Specifically, we encourage NTIA to center transparency and documentation requirements that focus on sociotechnical impacts, the need for federal action and regulation grounded by the AI Bill of Rights, and the need for further sociotechnical research on AI systems and their impacts as it advises the President on building a robust AI accountability ecosystem.

Respectfully submitted,

Serena Oduro, Senior Policy Analyst
Ranjit Singh, Researcher, AI on the Ground
Jacob Metcalf, Program Director, AI on the Ground

Endnotes

[1] Our comment also draws heavily from our research publications. See: Jacob Metcalf, Emanuel Moss, Ranjit Singh, Emnet Tafesse, and Elizabeth Anne Watkins. 2023. Taking Algorithms to Courts: A Relational Approach to Algorithmic Accountability. In 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’23), June 12–15, 2023, Chicago, IL, USA. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3593013.3594092 and Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, Madeleine Clare Elish, and Jacob Metcalf. (June 29, 2021). Assembling Accountability: Algorithmic Impact Assessment for the Public Interest. Available at: https://datasociety.net/wp-content/uploads/2021/06/Assembling-Accountability.pdf.

[2] Selbst, Andrew D. “An Institutional View of Algorithmic Impact Assessments.” Harv. J.L. & Tech. 35 (2021): 6.

[3] ibid.

[4] Balkin, Jack M. “The Three Laws of Robotics in the Age of Big Data.” Faculty Scholarship Series 5159 (2017). https://digitalcommons.law.yale.edu/fss_papers/5159.

[5] Tutt, Andrew. “An FDA for Algorithms.” Administrative Law Review 69, no. 1 (2017): 83–123. https://doi.org/10.2139/ssrn.2747994.

[6] Lemay, Mathieu. “Understanding Canada’s Algorithmic Impact Assessment Tool.” Toward Data Science (blog), June 11, 2019. https://towardsdatascience.com/understanding-canadas-algorithmic-impact-assessment-tool-cd0d3c8cafab.

[7] Council of Europe, and European Parliament. “Regulation on European Approach for Artificial Intelligence Laying Down a Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts,” 2021. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence-artificial-intelligence.

[8] https://www.washingtonpost.com/technology/2023/03/14/snapchat-myai/

[9] http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf

[10] https://themarkup.org/newsletter/hello-world/the-seven-year-struggle-to-hold-an-out-of-control-algorithm-to-account

[11] Danielle Keats Citron, Technological Due Process, 85 WASH. U. L. REV. 1249 (2008). Available at: https://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2

[12] The five principles are safe and effective systems, freedom from algorithmic discrimination, data privacy, notice and explanation, and human alternatives, consideration, and fallback. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[13] Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. "Datasheets for datasets." Communications of the ACM 64, no. 12 (2021): 86 – 92.

[14] Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. "Model cards for model reporting." In Proceedings of the conference on fairness, accountability, and transparency, pp. 220 – 229. 2019.

[15] Recent layoffs of teams critical to technology companies' responsible AI efforts also highlight the need for the government to require accountable AI practices. https://www.theverge.com/2023/3/13/23638823/microsoft-ethics-society-team-responsible-ai-layoffs