

Hon. Alan Davidson
Assistant Secretary of Commerce for Communications and Information
National Telecommunications and Information Administration
U.S. Department of Commerce
1401 Constitution Ave N.W.
Washington, DC 20230

Re: AI Accountability Policy Request for Comment

Introduction:

To ensure that artificial intelligence development remains aligned with our country's values and interests, it is critical that we establish a robust regulatory foundation that is rooted in trust and transparency.

Question – *A.I. accountability measures have been proposed in connection with many different goals, including those listed below. To what extent are there tradeoffs among these goals? To what extent can these inquiries be conducted by a single team or instrument? A.) The A.I. system does not substantially contribute to harmful discrimination against people.*

According to the National Institute of Standards and Technology, biases are an intrinsic aspect of artificial intelligence and exist in three distinct forms: statistical and computational bias, systemic bias, and human bias.¹ To ensure that A.I. systems do not substantially contribute to discrimination, regulators should implement accountability measures that focus on achievable goals through transparency standards and human oversight rather than the unattainable target of eliminating bias altogether.

Statistical and computational biases derive from the failure of a dataset to represent a full population.² This form of bias originates from systematic rather than random errors and can manifest even in the absence of any overt prejudiced, biased, or discriminatory intent.³ These inherent biases can undetectably permeate datasets and the algorithms that are utilized within the system.⁴ For example, data can become misrepresented by oversimplified mathematical representations, inadequate data cleaning, or incorrect treatment of outliers.⁵ While it may be tempting for regulators to try to eliminate this form of bias, the datasets used to train A.I. pose a unique challenge because of their large size. For example, GPT-3, product of OpenAI, uses a 175 billion parameter Large Language Model.⁶

A.I. systems also include systemic biases, often unrelated to conscious prejudice or discrimination, which arise from adherence to prevailing rules and norms. These biases, which can mirror the prevailing standards within office environments or cultural contexts, have the potential to shape the effectiveness

¹ National Institute of Standards and Technology, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, March 2022, p. 8, <https://doi.org/10.6028/NIST.SP.1270> (accessed June 5, 2023).

² *Ibid.*, p. 9.

³ *Ibid.*

⁴ *Ibid.*

⁵ *Ibid.*

⁶ Cameron R. Wolfe, "Language Model Scaling Laws and GPT-3," *Towards Data Science*, December 9, 2022, <https://towardsdatascience.com/language-model-scaling-laws-and-gpt-3-5cdc034e67bb> (accessed June 5, 2023).

and performance of an A.I. system.⁷ Systemic biases can be difficult to detect, and regulators will likely struggle to solve for the full scope of such biases.⁸

As NIST describes in their publication, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, human biases in A.I. systems stem from “systematic errors in human thought based on a limited number of heuristic principles and predicting values to simpler judgmental operations.”⁹ These biases are implicit and affect how individuals and groups interpret information, make decisions, and fill gaps in knowledge, including when interacting with A.I. outputs. They persist throughout the entire A.I. lifecycle, from development to deployment, in institutional, group, and individual decision-making. Even if all statistical, computational, and systemic biases were able to be removed from A.I. systems, pervasive and often undetectable human biases would be difficult to constrain. Cognitive and perceptual biases, which are inherent to human thinking, span different areas and are not exclusive to our interactions with A.I. They shape our understanding and influence our perceptions.¹⁰ Subsequently, human biases will also be present within any regulatory framework used to eradicate bias or to protect against discrimination. In other words, human bias is unavoidable.

When considering rules to mitigate bias, regulators should prioritize A.I. explainability, meaning the ability for humans to understand an artificial intelligence system’s behavior.¹¹ This measure is particularly important when it comes to maintaining oversight over deep neural networks, as their decision-making processes are shaped by prior experiences or training data.¹² While deep neural networks may successfully accomplish their assigned tasks, the “black box” nature of their decision-making makes it difficult for users or regulators to grasp how outcomes are determined.¹³ To facilitate a transparent environment, regulators should release these models and require organizations developing these models to make their code open-source wherever possible, like that of OpenAI’s GPT-2 in 2019.¹⁴ By doing so, both regulators and users will gain deeper insights into the inner workings of these systems, their developmental trajectory, and their decision-making processes, thereby establishing robust checks and balances for A.I. accountability.¹⁵

Through enacting these A.I. explainability measures, regulators will help preserve human oversight, and provide the transparency necessary for regulators to expedite the identification and rectification of biases.

Question – *What aspects of human rights and/or industry Environmental, Social, and Governance (ESG) assurance systems can and should be adopted for A.I. accountability?*

⁷ Ibid, p. 6.

⁸ Greg Satell and Yassmin Abdel-Magied, “A.I. Fairness Isn’t Just an Ethical Issue,” *Harvard Business Review*, October 20, 2020, <https://hbr.org/2020/10/ai-fairness-isnt-just-an-ethical-issue> (accessed June 7, 2023).

⁹ National Institute of Standards and Technology, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, p. 9.

¹⁰ Ibid.

¹¹ Michael Horowitz and Paul Scharre, “Artificial Intelligence: What Every Policymaker Needs to Know,” Center for a New American Security, June 19, 2018, <https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know> (accessed June 8, 2023).

¹² Ibid.

¹³ Ibid.

¹⁴ Miles Brundage et al., “Better Language Models and Their Implications,” OpenAI, February 14, 2019, <https://openai.com/research/better-language-models> (accessed June 12, 2023).

¹⁵ Ibid.

Environmental, Social, and Governance (ESG) standards are subjective and unsuitable for A.I. accountability. To uphold impartiality in A.I. systems, regulators should prioritize constructing an objective and unbiased framework that tackles the array of concerns associated with A.I. Such an approach would pave the path towards a more reliable and all-encompassing A.I. accountability framework, ensuring that the challenges posed by A.I. are effectively addressed.

As highlighted in a report published by the American Council for Capital Formation, when it comes to ESG compliance, Environmental and Social disclosures lack standardized rules and guidelines, and there is no established auditing process to validate reported data. Consequently, agencies are forced to rely on assumptions when assessing such information.¹⁶ If A.I. accountability standards are contingent upon such assumptions, the enforcement of accountability measures will be susceptible to subjective interpretation and implementation. The very process of determining which ESG factors to prioritize creates a potential for ideological bias, as certain priorities may be favored over others. In the case of Clearview A.I., an audit conducted using the ACLU's facial recognition accuracy methodology assessed that their data was "100% accurate across all demographic groups." However, the audit failed to address how the facial recognition data was collected, which posed serious ethical and legal dilemmas,¹⁷ highlighting the complexity of A.I. system auditing.¹⁸ As demonstrated through the Clearview A.I. case, without well-defined standards governing the audit process, assessments possess little value.¹⁹ Additionally, the absence of universally established regulations governing ESG compliance renders an accountability framework reliant on such standards ill-equipped to fulfill its intended objectives.

Furthermore, according to the European Corporate Governance Institute, the decision-making process regarding ESG rating criteria exhibits a notable degree of subjectivity. The report also outlines that ESG ratings agencies rely heavily on the information provided by the companies under evaluation.²⁰ This dynamic is particularly concerning considering that companies falling short of earnings expectations tend to be more likely to mention ESG performance criteria than firms that performed well on objective, shareholder-based measures.²¹ If regulators introduced a similar ESG-based accountability framework for A.I. systems, it is conceivable that underperforming or unscrupulous entities may exploit the system by strategically flaunting their ESG compliance to evade regulatory scrutiny or to receive special treatment. The possibility of such an occurrence further undermines the neutrality that should be inherent in the A.I. decision-making processes. By incorporating ESG, we risk not only perpetuating

¹⁶ Timothy M. Doyle, "Ratings That Don't Rate: The Subjective World of ESG Rating Agencies," American Council for Capital Formation, July 2018, https://accfcorp.gov/wp-content/uploads/2018/07/ACCF_RatingsESGReport.pdf (accessed June 5, 2023).

¹⁷ Inioluwa Deborah Raji et al., "Outsider Oversight: Designing a Third Party Audit Ecosystem for A.I. Governance," *Proceedings of Conference on A.I., Ethics, and Society*, July 2022, <https://doi.org/10.1145/3514094.3534181> (accessed June 6, 2023).

¹⁸ Richard Benjamins, Alberto Barbado, and Daniel Sierra, "Responsible A.I. by Design in Practice," Telefónica, December 20, 2019, p. 5, <https://arxiv.org/ftp/arxiv/papers/1909/1909.12838.pdf> (accessed June 7, 2023).

¹⁹ Inioluwa Deborah Raji et al., "Outsider Oversight: Designing a Third Party Audit Ecosystem for A.I. Governance," *Proceedings of Conference on A.I., Ethics, and Society*, July 2022, <https://doi.org/10.1145/3514094.3534181> (accessed June 6, 2023).

²⁰ Florian Berg, Kornelia Fabisik, and Zacharias Sautner, "Is History Repeating Itself? The (Un)Predictable Past of ESG Ratings," working paper, European Corporate Governance Center, August 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3722087 (accessed June 5, 2023).

²¹ Ryan Flugum and Matthew E. Souther, "Stakeholder Value: A Convenient Excuse for Underperforming Managers?" March 29, 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3725828 (accessed June 5, 2023).

systemic and human biases within A.I. systems, but also hindering the progress made in addressing these issues.

If regulators are interested in promoting fairness and impartiality within A.I. systems, then they should construct an objective and unbiased framework for A.I. accountability free of ESG requirements. The framework should be firmly grounded in human-centered values and embrace fundamental freedoms. By requiring companies to adhere to these established and accepted principles, regulators can eliminate the biases that would be present in an ESG-based accountability framework. This approach will ensure that accountability protocols will uphold A.I. accountability for all systems.

Question – *As governments at all levels increase their use of A.I. systems, what should the public expect in terms of audits and assessments of A.I. systems deployed as part of public programs? Should the accountability practices for A.I. systems deployed in the public sector differ from those used for private sector A.I.? How can government procurement practices help create a productive A.I. accountability ecosystem?*

As governments and consumers increase their use of A.I. systems, the public should expect regulators to establish a robust transparency system that includes frequent audits and assessments of these systems. Regulators should prioritize the implementation of rulesets that mandate A.I. explainability and open data access to ensure trust and safety in both public and private domains. Such rulesets should be applicable to all A.I. systems that engage with the general public or consumers so that regulators can ensure firms are held accountable for their products, and that A.I. systems are safely deployed and used.

Independent audits play a critical role in the governance of artificial intelligence due to their ability to assess the integrity and fairness of A.I. systems. Regulators should implement an auditing framework that includes risk assessments prior to deploying highly automated systems, an auditable trail for A.I. decision-making that enables an assessment of how an A.I. system arrived at its decision, and a way to assess a given system's adherence to regulatory requirements so that they can guarantee an impartial and comprehensive evaluation of these systems.²² The significance of these independent audits becomes even more pronounced in the context of DNNs, which can contain tens of billions of parameters,²³ providing ample opportunity for biases to be concealed within an A.I. system.²⁴ Algorithmic impact assessments are also necessary within corporate contexts, where frequent assessments of design, privacy, and biases are all vital to ensuring informed applications of artificial intelligence systems.²⁵ Therefore, incorporating independent audits as a fundamental aspect of A.I. governance is essential to upholding the integrity and fairness of A.I. in the public and private spheres alike.

Furthermore, clear explanations for A.I. decision-making are vital for maintaining trust and understanding among stakeholders and to mitigate potential risks associated with opaque A.I. systems

²² Gregory Falco et al. "Governing A.I. Safety Through Independent Audits," *Nature Machine Intelligence*, Vol. 3 (July 2022), pp. 566-570, https://www.nature.com/articles/s42256-021-00370-7.epdf?sharing_token=pEZIDGQiqq9k_YBp8GUq4NRgN0jAjWel9jnR3ZoTv0OgQK3tLTRjd7yOOKuLlVicxyGXDcoZKa9GYzX01rLfem0w5iR-ikMTmV6sLu9HwFyxINcFmR6kHxTO0hD2pNh8ggqPVsrLjCP_ozS6M93iW8X_nFWU1V21ff0_uV0sJmc%3D (accessed June 8, 2023).

²³ Richard Baraniuk, David Donoho, and Matan Gavish, "The Science of Deep Learning," *Proceedings of the National Academic of Sciences*, Vol. 117, No. 48 (November 23, 2020), <https://www.pnas.org/doi/epdf/10.1073/pnas.2020596117> (accessed June 7, 2023).

²⁴ Richard Benjamins, Alberto Barbado, and Daniel Sierra, "Responsible A.I. by Design in Practice," p. 4.

²⁵ Inioluwa Deborah Raji et al., "Outsider Oversight: Designing a Third Party Audit Ecosystem for A.I. Governance," p. 558.

that are or will be used in the public sphere.²⁶ Regulators should pursue rules that require A.I. explainability.²⁷ While this may seem simple, as A.I. systems become more sophisticated, the task of tracing their decision-making process becomes increasingly challenging. The process through which A.I. systems gain their intelligence involves continuous ingestion of data, evaluation of the predictive capabilities of various algorithmic combinations, and iterative refinement of the resulting model, all of which occur at extraordinary speeds. Therefore, regulators should pursue rules that are centered around explainability and that facilitate the adoption of standardized development and testing practices. Through this approach, regulators can guarantee the replicability of A.I. system outcomes while effectively mapping, quantifying, and mitigating associated risks in a consistent manner.²⁸

Overall, as regulators consider rules aimed at holding A.I. systems accountable, they must ensure that A.I. technologies do not operate in secret. Regulators should pursue rules that require independent audits of A.I. systems and clear explanations of how A.I. systems make decisions. This will allow society to understand and navigate the significant impacts of these technologies. This commitment to transparency not only improves accountability but also enables public scrutiny.

Respectfully Submitted,

Jake Denton

²⁶ Richard Benjamins, Alberto Barbado, Daniel Sierra, "Responsible A.I. by Design in Practice," p. 5.

²⁷ Paul Scharre and Michael Horowitz, "Artificial Intelligence: What Every Policymaker Needs to Know."

²⁸ National Institute of Standards and Technology, *Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*, p. 43.