



## **Human Rights, Transparency, and Stakeholder Involvement in AI Accountability**

A comment submitted to the National Telecommunications and Information Administration

### **Authors:**

Anna-Maria Gueorguieva, Sinuhe Cruz, Emma Kallina, & Caitlin-Kraft Buchman

### ***Executive Summary***

This comment addresses questions of transparency, human rights, and stakeholder involvement put forth by the National Telecommunications and Information Administration. We urge the NTIA to consider AI accountability through a framework of human rights. First, we address transparency & accountability through the perspective of due process. Then, we provide information on the “right to know”, urging that AI accountability measures include the requirement that individuals should be notified when an automated decision has been made about them. Finally, we conclude with information and suggestions for stakeholder involvement, a part of AI accountability that we believe to be unwaveringly necessary.

## I. Introduction

### *The Necessity for General Law / Regulation*

The current lack of any AI regulation in the US makes it difficult to implement value-based guidelines, such as Trustworthy AI principles, in practice<sup>1</sup>. For example, a team aiming to develop AI technology for the National AI Institute within the Department of Veterans Affairs stated they adhered to Trustworthy AI principles (in reference to Executive Order 13960), but “admit[ed] they do not know the NAII’s actual litmus test for applying these principles”<sup>2</sup>. This means that those producing AI for government and private sectors are not uniform in their understanding of value-based principles. This is dangerous, as the potential for harm through AI decision making is high. Thus, addressing *Question 30*, we urge that regulation be passed that translates abstract principles into required action items, such as: technical audits, algorithmic notification usage, requirement of stakeholder involvement, and impact assessments. This comment further discusses the “right to know” presented via algorithmic notification usage and stakeholder impact further in order to provide normative ideas for what regulation should include.

### *Human Rights Obligations & Frameworks*

Addressing *Question 13* and *Question 1d*, we believe that *all* aspects of human rights should be adopted for AI accountability, and that AI audits and assessments should promote the goals of human rights. The United States has ratified the International Covenant on Civil and Political Rights and the International Convention on the Elimination of All Forms of Racial Discrimination, and has signed multiple other international human rights treaties. Thus, the United States should (and is legally required to) incorporate human rights into their accountability measures. We urge the

---

<sup>1</sup> Anna Gueorguieva, *Artificial Intelligence & Regulatory Actors in the Sociotechnical System of the Department of Veterans Affairs*. 2023

<sup>2</sup> Department of Veterans Affairs. *NAII Announces the ASPIRE AI Tech Sprint 2022 Winners – NAII AI Tech Sprints Recorded video of live observation on February 2022: [Link](#)* (Accessed April 28, 2023 ).

National Telecommunications and Information Administration to value and uphold signed and ratified international human rights treaties, and use human rights frameworks and language in creating AI accountability measures, directives, and regulations.

## II. Towards Accountable AI: Addressing the Lack of Transparency

*Why do transparency and explainability matter?*

Decisions with a significant impact on individual and collective rights have become increasingly mediated by opaque algorithmic systems, without the affected persons receiving adequate and sufficient information about the logic, criteria, and the elements under which such decisions are made. Addressing *Question 3e*, we believe that transparency and explainability mechanisms play an important role in guaranteeing the information self-determination of individuals subjected to automated decision-making, enabling them to access and understand the output decision and its underlying elements, and thus providing pathways for those who wish to challenge and request a review of the decision. From this perspective, enabling AI-based systems with adequate transparency and explanation to affected people about their uses, capabilities, and limitations amounts to applying the due process safeguards derived from constitutional law in the analogue world to the digital world. That is, in the same way that State interference in the rights of a citizen must be proportionate, motivated, informed, and well-founded, providing the individual with the possibility of defending themselves and contesting the decision, automated decisions must also be subject to the same type of procedural safeguards and guarantees.

Translated to the digital world, the due process general clause can lead to what has been described as ‘informational due process’<sup>3</sup> or ‘technological due process’<sup>4</sup>, in which transparency and

---

<sup>3</sup> Kate Crawford & Jason Schultz. *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*. Rochester, NY: Social Science Research Network, 2014. Available at: <https://papers.ssrn.com/abstract=2325784>; HILDEBRANDT, Mireille; VRIES, Katja de. *Privacy, Due Process and The Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*. 1st edition. London: Routledge, 2013.

<sup>4</sup> Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008). Available at: [https://openscholarship.wustl.edu/law\\_lawreview/vol85/iss6/2](https://openscholarship.wustl.edu/law_lawreview/vol85/iss6/2).

explainability play a pivotal role. Within this framework, transparency and explainability help to guarantee that automated decision-making made by AI is (i) impartial (free of biases that could lead to harmful discrimination), (ii) informed (providing the affected people with the necessary and meaningful information regarding the decision-making), (iii) explainable (enabling people to effectively understand the fundamentals and the output of the decision); (iv) appealable (subjected to be challenged and contested by the affected individuals) and (v) subjected to revision made by a neutral and impartial source, preferably carried out by a natural person.

*Transparency and explainability of AI-based systems: the need for a contextual and systemic approach*

Addressing *Question 27*, an effective policy aimed at promoting AI accountability must take into account that there are different kinds of opacity that can derive from AI systems, such as opacity originating from intentional corporate or state secrecy, opacity originating from technical illiteracy, and opacity that arises from the characteristics of complex and unsupervised algorithms<sup>5</sup>, each one requiring different transparency and explainability measures. There is no ‘one-size-fits-all’ approach and this should be taken into account in policymaking.

NTIA's AI Accountability Policy should also consider establishing a contextual approach to its transparency and explainability mechanisms. There are different types of explanations, each of which is more or less adequate depending on the context, the recipients, the stage of the AI lifecycle, and the intended goals.<sup>6</sup> The transparency and accountability requirements should also be tailored and calibrated according to the amount of risk presented by the specific sector or domain in which the AI system is being deployed (e.g. health, immigration, public security, real-time remote biometric identification, law enforcement, etc.) or the type and characteristics of the AI model (e.g. supervised or unsupervised). Finally, when considering this contextual approach, it is essential to put

---

<sup>5</sup> Jenna Burrell. How the Machine 'Thinks:' Understanding Opacity in Machine Learning Algorithms (September 15, 2015). Available at SSRN: <https://ssrn.com/abstract=2660674> or <http://dx.doi.org/10.2139/ssrn.2660674>

<sup>6</sup> Brkan, M., & Bonnet, G. (2020). Legal and Technical Feasibility of the GDPR's Quest for Explanation of Algorithmic Decisions: Of Black Boxes, White Boxes and Fata Morganas. *European Journal of Risk Regulation*, 11(1), 18-50. doi:10.1017/err.2020.10

the recipients and the affected individuals at the center of the transparency and explainability framework, since different stakeholders will have different necessities, different digital literacy and technical expertise levels, and different expectations regarding transparency and explanations.<sup>7</sup> “Opening up the black box” to reveal code may not produce the understanding we need of the environments the code enables, and what impact the code may have and has on human beings. It is necessary to articulate the information provided through effective explanation in order to enable the affected individual to understand the potential negative effects and limitations of AI systems, then turning this information into actionable knowledge for all the concerned stakeholders.

### **III. The “Right to Know”: A Call for Algorithmic Usage Notification**

*The “right to know” aims to establish the basic principle and requirement that individuals have a right to be informed when automated systems, algorithms, and AI have made a decision about them. The concept of the right to know was initially written for a proposed resolution to the United Nations Human Rights Council by Caitlin Kraft-Buchman and Women at the Table..*

By informing individuals when they are subject to algorithmic decisions, it empowers them to question, challenge, and seek redress for any potential biases, inaccuracies, or injustices that may result from the use of such systems. Furthermore, the right to know enables individuals to make informed choices about their participation in digital spaces, fostering a more equitable and democratic society. This right is closely intertwined with the concepts of transparency and accountability discussed in section II, especially when it comes to upholding due process. Although there isn’t a specific human right called the right to know enshrined in international human rights law, the concept of the right to know is closely related to other established human rights, such as the right to freedom of expression and the right to access information. These rights are recognized

---

<sup>7</sup> Information Commissioner’s Office; The Alan Turing Institute. Explaining decisions made with AI. Information Commissioner’s Office. Available at: <https://ico.org.uk/media/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/explaining-decisions-made-with-artificial-intelligence-1-0.pdf>.

in international legal instruments, such as the Universal Declaration of Human Rights (UDHR) and the International Covenant on Civil and Political Rights (ICCPR). The right to access information is an essential aspect of the right to know. This right entitles individuals to seek, receive, and impart information, including information held by public authorities. The right to access information is vital for promoting transparency, government accountability, and public participation in decision-making processes. In the context of the digital age and the rise of artificial intelligence and algorithmic decision-making, the concept of the right to know is evolving to encompass the need for individuals to be informed when they are subject to decisions made by algorithms or automated systems.

*The right to know when an individual has been affected or touched by algorithmic decision-making includes, but is not limited to, the right to:*

1. Be informed when a decision concerning them has been made or facilitated by an algorithmic system;
2. Access information about the purposes, functioning, and significant characteristics of the algorithmic system used in the decision-making process;
3. Obtain an explanation of the decision and the logic involved, including any significant factors, inputs, or correlations that influenced the decision, as well as why the algorithm was used;
4. Request human intervention or review of the decision, especially when the decision has significant consequences for their lives or interests; and
5. Challenge or appeal the decision, including the right to contest the algorithmic system's legality, fairness, and accuracy, and to seek appropriate remedies.

#### **IV. Call for Enforcing Stakeholder Involvement**

*Academic research in the trustworthy AI space is increasingly nominating stakeholder involvement as the missing link between theoretical considerations and their practical implementation. Thus, it is essential that stakeholder involvement forms a substantial part of AI accountability.*

To translate the highly abstract concepts of trustworthy AI (such as safety, fairness, privacy, notice and explanation) into actionable requirements for a specific AI system, it is essential to involve the system's stakeholders. Everyone who is affected by an AI system's development, deployment, or operation is considered a stakeholder of the system. Involving them in the system's design would follow the credo: all communities that are expected to be influenced by an AI system should substantially inform its design. In practice, this means that the stakeholders should advise on the concrete manifestation of the different aspects of trustworthy AI in the system, i.e. their importance, hierarchy, and the system features that would fulfill them. For example, for 'notice and explanation' the stakeholders should inform the depth, form, and timings of the provided information. Thus, trustworthy AI is not only a system that encapsulates a broad set of technical and socio-technical attributes, but that involves its stakeholders in defining what exactly these attributes mean in the context of system application.

A second substantial benefit of stakeholder involvement is the increased detection of a system's unintended effects through an increased understanding of a system's socio-technical context. Consulting the communities present in the system's application context (including marginalized groups) will provide insights into their diverse experiences. This enables a better understanding of existing harms that might be exacerbated by the system, as well as new risks that might arise (i.e. reduce uncertainty). As a consequence, stakeholder involvement should be required by AI regulation.

This relates to the AI Accountability Policy in the following ways:

1. **Stakeholder involvement can act as a method of internal, informal auditing.** Through testing the system's (planned) design against the needs of its stakeholders, stakeholder involvement can deliver insights that can inform decisions regarding the system's features, its point of deployment, required updates, or when its retirement is necessary. An increased awareness of the system's stakeholders' needs and values, paired with the explicit ownership for the design decisions that resulted from these, seems to be a promising way to increase the feeling of responsibility in AI practitioners. Thus, stakeholder involvement might promote responsible system design without the explicit legal obligation to do so. (*Relating to question 1b, 1e, and 16a*).
  - a. **Note regarding question 16a:** AI accountability mechanisms should take an organization's effort to develop an AI system responsibly into account. These efforts should include stakeholder involvement for a more detailed understanding of the socio-technical context in which the system is / will be deployed. This understanding should then inform the more narrow, technical aspects. In other words, the socio-technical, broader perspective is required to make informed choices regarding technical aspects. Thus, stakeholder involvement should be a substantial part of AI accountability.
2. **Stakeholder involvement should be included as a sub-point of question 3.**
  - a. Proposed phrase: "There has been adequate consultation with the relevant stakeholders of the system, and the gained insights substantially, and in an auditable manner, influenced the system's design and risk management."



- b. Thereby, it is essential that stakeholder involvement is conducted throughout the AI pipeline, including after deployment. This ensures that the system's design is still meeting the requirements of the changing needs of its context. *(Relating to question 2, 3, and 15a).*
3. **Stakeholder involvement can reduce the uncertainty regarding a system's harms.** AI systems imply a certain level of uncertainty and their deployment might trigger unpredicted consequences. This hampers AI accountability since organizations might argue that they were not able to foresee a harm that their system caused. Adequate stakeholder involvement (i.e. involving a diverse and broad set of stakeholders and acting based on the gained insights) could act as a proof for a company's actual effort to foresee potential harms.
  - a. This would require **in-depth documentation**, including: Justification of why a specific set of stakeholders was selected (and other stakeholders were not), when they were involved and for what purpose, the gained insights and method of a point of involvement, as well as the actions taken as a reaction to these insights and a justification of this choice. The responsibility for every action and decision should be clearly assigned. *(Relating to question 2, 4, 15b, 15d, 17, and 21)*
  - b. **Note on question 17:** The risk of an AI system should be substantially informed by socio-technical considerations, which in turn are substantially informed through insights gathered from the system's stakeholders. Thus, a system's stakeholders considerably influence the risk assigned to it.

We hope that the above considerations will lead to the increased advocacy for stakeholder involvement in AI regulation as well as AI development practice.

## V. About <AI & Equality> and Comment Authors

<AI & Equality> is an initiative of Women at the Table. We are a global community of diverse individuals coming together from many different backgrounds, disciplines, and expertise to call for a human-rights led AI future. The authors of this piece hold degrees in multiple legal, computer science, and data science fields with experience working in AI regulation, stakeholder involvement design, and AI development. Coming from these backgrounds and experiences, we felt called to urge the NTIA to consider how human rights frameworks should be important in AI accountability. It is especially important that the NTIA consider the legal obligation to adhere to international human rights treaties and United States due process laws when creating AI accountability policy. We call for: transparency and explainability through actual stakeholder understanding of AI, the requirement of algorithmic usage notification (or the “right to know”), and for active stakeholder engagement in AI development.

We thank you for your time and hope that this comment provides useful insights and guiding ideas for AI accountability policy development.

Sincerely,

*Anna-Maria Gueorguieva, Sinuhe Cruz, Emma Kallina, and Caitlin-Kraft Buchman*



Signed June 12, 2023

## Appendix of NTIA Questions Referenced

**Question 1b:** What are assessments or internal audits most useful for? What are external assessments or audits most useful for?

**Question 1d:** Should AI audits or assessments be folded into other accountability mechanisms that focus on such goals as human rights, privacy protection, security, and diversity, equity, inclusion, and access? Are there benchmarks for these other accountability mechanisms that should inform AI accountability measures?

**Question 1e:** Can AI accountability practices have meaningful impact in the absence of legal standards and enforceable risk thresholds? What is the role for courts, legislatures, and rulemaking bodies?

**Question 2:** Is the value of certifications, audits, and assessments mostly to promote trust for external stakeholders or is it to change internal processes? How might the answer influence policy design?

**Question 3:** AI 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?

**Question 3e:** There has been adequate transparency and explanation to affected people about the uses, capabilities, and limitations of the AI system.

**Question 4:** Can AI accountability mechanisms effectively deal with systemic and/or collective risks of harm, for example, with respect to worker and workplace health and safety, the health and safety of marginalized communities, the democratic process, human autonomy, or emergent risks?

**Question 13:** What aspects of human rights and/or industry Environmental, Social, and Governance (ESG) assurance systems can and should be adopted for AI accountability?

**Question 15a:** Where in the value chain should accountability efforts focus?

**Question 15b:** How can accountability efforts at different points in the value chain best be coordinated and communicated?

**Question 15d:** Since the effects and performance of an AI system will depend on the context in which it is deployed, how can accountability measures accommodate unknowns about ultimate downstream implementation?

**Question 16a:** Should AI accountability mechanisms focus narrowly on the technical characteristics of a defined model and relevant data? Or should they feature other aspects of the socio- technical

system, including the system in which the AI is embedded? When is the narrower scope better and when is the broader better? How can the scope and limitations of the accountability mechanism be effectively communicated to outside stakeholders?

**Question 17:** How should AI accountability measures be scoped (whether voluntary or mandatory) depending on the risk of the technology and/or of the deployment context? If so, how should risk be calculated and by whom?

**Question 21:** What are the obstacles to the flow of information necessary for AI accountability either within an organization or to outside examiners? What policies might ease researcher and other third-party access to inputs necessary to conduct AI audits or assessments?

**Question 27:** What is the role of intellectual property rights, terms of service, contractual obligations, or other legal entitlements in fostering or impeding a robust AI accountability ecosystem? For example, do nondisclosure agreements or trade secret protections impede the assessment or audit of AI systems and processes? If so, what legal or policy developments are needed to ensure an effective accountability framework?

**Question 30:** What role should government policy have, if any, in the AI accountability ecosystem? For example: a. Should AI accountability policies and/or regulation be sectoral or horizontal, or some combination of the two?