Billing Code 3510-60-P

DEPARTMENT OF COMMERCE

National Telecommunications and Information Administration

[Docket No. 230407-0093]

RIN 0660-XC057

AI Accountability Policy Request for Comment

AGENCY: National Telecommunications and Information Administration, U.S. Department of Commerce.

ACTION: Notice, Request for Comment.

SUMMARY: The National Telecommunications and Information Administration (NTIA) hereby requests comments on Artificial Intelligence ("AI") system accountability measures and policies. This request focuses on self-regulatory, regulatory, and other measures and policies that are designed to provide reliable evidence to external stakeholders – that is, to provide assurance -- that AI systems are legal, effective, ethical, safe, and otherwise trustworthy. NTIA will rely on these comments, along with other public engagements on this topic, to draft and issue a report on AI accountability policy development, focusing especially on the AI assurance ecosystem.

DATES: Written comments must be received on or before June 10, 2023.

ADDRESSES: All electronic public comments on this action, identified by Regulations.gov docket number NTIA-2023-0005, may be submitted through the Federal e-Rulemaking Portal at www.regulations.gov. The docket established for this request for comment can be found at www.regulations.gov, NTIA-2023-0005. Click the "Comment Now!" icon, complete the required fields, and enter or attach your comments. Additional instructions can be found in the "Instructions" section below after "Supplementary Information."

FOR FURTHER INFORMATION CONTACT: Please direct questions regarding this Notice to Travis Hall at thall@ntia.gov with "AI Accountability Policy Request for Comment" in the subject line, or if by mail, addressed to Travis Hall, National Telecommunications and Information Administration, U.S. Department of Commerce, 1401 Constitution Avenue, NW, Room 4725, Washington, DC 20230; telephone: (202) 482-3522. Please direct media inquiries to NTIA's Office of Public Affairs, telephone: (202) 482–7002; email: press@ntia.gov.

SUPPLEMENTARY INFORMATION:

Background and Authority:

Advancing trustworthy Artificial Intelligence ("AI") is an important federal objective.[1] The National AI Initiative Act of 2020[2] established federal priorities for AI, creating the National AI Initiative Office to coordinate federal efforts to advance trustworthy AI applications, research, and U.S. leadership in the development and use of trustworthy AI in the public and private sectors.[3] Other legislation, such as the landmark CHIPS and Science Act of 2022, also support the advancement of trustworthy AI.[4] These initiatives are in accord with Administration efforts to advance American values and leadership in AI[5] and technology platform accountability[6] and to promote "trustworthy artificial intelligence" as part of a national security

[1] *See generally*, Laurie A Harris, Artificial Intelligence: Background, Selected Issues, and Policy Considerations, CRS 46795, U.S. Library of Congress: Congressional Research Service, (May 19, 2021), at 16-26, 41-42, https://crsreports.congress.gov/product/pdf/R/R46795 (last visited Feb. 1, 2023).
[2] The National Artificial Intelligence Initiative Act of 2020, Pub. L. 116-283, 134 Stat. 3388 (Jan. 1, 2021).
[3] U.S. National Artificial Intelligence Initiative Office, Advancing Trustworthy AI Initiative, https://www.ai.gov/strategic-pillars/advancing-trustworthy-ai (last visited Jan. 19, 2023).
[4] *See, e.g.,* CHIPS and Science Act of 2022, Pub. L. No. 117-167, 136 Stat. 1392 (Aug. 9, 2022) (providing support and guidance for the development of safe, secure, and trustworthy AI systems, including considerations of fairness and bias as well as the ethical, legal, and societal implications of AI more generally).
[5] *Supra* note 2 (implemented though the National Artificial Intelligence Initiative, https://ai.gov (last visited Jan. 19, 2023)).
[6] White House, Readout of White House Listening Session on Tech Platform Accountability (Sept. 8, 2022) [Tech Platform Accountability], https://www.whitehouse.gov/briefing-room/statements-releases/2022/09/08/readout-of-white-house-listening-session-on-tech-platform-accountability (last visited Feb. 1, 2023).

strategy.[7] Endeavors that further AI system governance to combat harmful bias and promote equity and inclusion also support the Administration's agenda on racial equity and support for underserved communities.[8] Moreover, efforts to advance trustworthy AI are core to the work of the Department of Commerce. In recent public outreach, the International Trade Administration noted that the Department "is focused on solidifying U.S. leadership in emerging technologies, including AI" and that the "United States seeks to promote the development of innovative and trustworthy AI systems that respect human rights, [and] democratic values, and are designed to enhance privacy protections."[9]

To advance trustworthy AI, the White House Office of Science and Technology Policy produced a Blueprint for an AI Bill of Rights ("Blueprint"), providing guidance on "building and deploying automated systems that are aligned with democratic values and protect civil rights, civil liberties, and privacy."[10] The National Institute of Standards and Technology (NIST) produced an AI Risk Management Framework, which provides a voluntary process for managing

---

[7] White House, Biden-Harris Administration's National Security Strategy (Oct. 12, 2022) at 21, https://www.whitehouse.gov/wp-content/uploads/2022/10/Biden-Harris-Administrations-National-Security-Strategy-10.2022.pdf (last visited Feb. 1, 2023) (identifying "trusted artificial intelligence" and "trustworthy artificial intelligence" as priorities). *See also* U.S. Government Accountability Office; Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities, GAO-21-519SP (June 30, 2021) (proposing a framework for accountable AI around governance, data, performance, and monitoring).

[8] *See* Advancing Racial Equity and Support for Underserved Communities Through the Federal Government, Exec. Order No. 13985, 86 Fed. Reg. 7009 (Jan. 25, 2021) (revoking Exec. Order No. 13058); Further Advancing Racial Equity and Support for Underserved Communities Through the Federal Government, Exec. Order No. 14091, 88 Fed. Reg. 10825, 10827 (Feb. 16, 2023) (specifying a number of equity goals related to the use of AI, including the goal to "promote equity in science and root out bias in the design and use of new technologies, such as artificial intelligence.").

[9] International Trade Administration, Request for Comments on Artificial Intelligence Export Competitiveness, 87 Fed. Reg. 50288, 50288 (Oct. 17, 2022) ("ITA is broadly defining AI as both the goods and services that enable AI systems, such as data, algorithms and computing power, as well as AI-driven products across all industry verticals, such as autonomous vehicles, robotics and automation technology, medical devices and healthcare, security technology, and professional and business services, among others.").

[10] White House, Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People (Blueprint for AIBoR) (Oct. 2022), https://www.whitehouse.gov/ostp/ai-bill-of-rights.

a wide range of potential AI risks.[11] Both of these initiatives contemplate mechanisms to advance the trustworthiness of algorithmic technologies in particular contexts and practices.[12] Mechanisms such as measurements of AI system risks, impact assessments, and audits of AI system implementation against valid benchmarks and legal requirements, can build trust. They do so by helping to hold entities accountable for developing, using, and continuously improving the quality of AI products, thereby realizing the benefits of AI and reducing harms. These mechanisms can also incentivize organizations to invest in AI system governance and responsible AI products. Assurance that AI systems are trustworthy can assist with compliance efforts and help create marks of quality in the marketplace.

NTIA is the President's principal advisor on telecommunications and information policy issues. In this role, NTIA studies and develops policy on the impacts of information and communications technology on civil rights;[13] transparency in software components;[14] and the use of emerging digital technologies.[15] NTIA's statutory authority, its role in advancing sound

---

[11] National Institute for Standards and Technology, Artificial Intelligence Risk Management Framework 1.0 (AI RMF 1.0) (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. *See also* National Artificial Intelligence Research Resource Task Force, Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource (Jan. 2023), https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf (last visited Feb. 1, 2023) (presenting a roadmap to developing a widely accessible AI research cyberinfrastructure, including support for system auditing).

[12] *See, e.g.,* AI RMF 1.0, *supra* note 11 at 11 (graphically showing test, evaluation, verification, and validation (TEVV) processes, including assessment and audit, occur throughout an AI lifecycle); Blueprint for AIBoR, *supra* note 10 at 27-28 (referring to "independent" and "third party" audits, as well as "best practices" in audits and assessments to ensure high data quality and fair and effective AI systems). *See also* Tech Platform Accountability (Sept. 8, 2022) (including the goal of promoting transparency in platform algorithms and preventing discrimination in algorithmic decision-making).

[13] National Telecommunications and Information Administration, Data Privacy, Equity and Civil Rights Request for Comments, 88 Fed. Reg. 3714 (Jan. 20, 2023).

[14] National Telecommunications and Information Administration, Software Bill of Materials (Apr. 27, 2021), https://ntia.gov/page/software-bill-materials (last visited Feb. 1, 2023).

[15] *See, e.g.*, National Telecommunications and Information Administration, Spectrum Monitoring – Institute for Telecommunications Sciences, https://its.ntia.gov/research-topics/spectrum-management-r-d/spectrum-monitoring (last visited Feb. 1, 2023).

Internet, privacy, and digital equity policies, and its experience leading stakeholder engagement processes align with advancing sound policies for trustworthy AI generally and AI accountability policies in particular.

Definitions and Objectives

Real accountability can only be achieved when entities are held responsible for their decisions. A range of AI accountability processes and tools (*e.g.*, assessments and audits, governance policies, documentation and reporting, and testing and evaluation) can support this process by proving that an AI system is legal, effective, ethical, safe, and otherwise trustworthy – a function also known as providing AI assurance.

The term "trustworthy AI" is intended to encapsulate a broad set of technical and socio-technical attributes of AI systems such as safety, efficacy, fairness, privacy, notice and explanation, and availability of human alternatives. According to NIST, "trustworthy AI" systems are, among other things, "valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy-enhanced, and fair with their harmful bias managed."[16] Along the same lines, the Blueprint identifies a set of five principles and associated practices to help guide the design, use, and deployment of AI and other automated systems. These are: (1) safety and effectiveness, (2) algorithmic discrimination protections, (3) data privacy, (4) notice and explanation, and (5) human alternatives, consideration and fallback.[17] These principles align with the trustworthy AI principles propounded by the Organisation for Economic Co-operation and Development (OECD) in 2019, which 46 countries have now

[16] AI RMF 1.0, *supra* note 11.
[17] White House, Blueprint for AIBoR, *supra* note 10.

adopted.[18] Other formulations of principles for responsible or trustworthy AI containing all or some of the above-stated characteristics are contained in industry codes,[19] academic writing,[20] civil society codes,[21] guidance and frameworks from standards bodies,[22] and other governmental instruments.[23] AI assurance is the practical implementation of these principles in applied settings with adequate internal or external enforcement to provide for accountability.

Many entities already engage in accountability around cybersecurity, privacy, and other risks related to digital technologies. The selection of AI and other automated systems for

[18] Organisation for Economic Co-operation and Development (OECD), Recommendation of the Council on Artificial Intelligence (May 22, 2019), https://www.oecd.org/gov/pcsd/recommendation-on-policy-coherence-for-sustainable-development-eng.pdf (last visited Feb. 1, 2023) (AI systems should (1) drive inclusive growth, sustainable development and well-being; (2) be designed to respect the rule of law, human rights, democratic values, and diversity; (3) be transparent; (4) be robust, safe, and secure; (5) and be accountable).
[19] *See, e.g.*, Microsoft, Microsoft Responsible AI Standard Reference Guide Version 2.0 (June 2022), https://query.prod.cms.rt.microsoft.com/cms/api/am/binary/RE4ZPmV (last visited Feb. 1, 2023) (identifying accountability, transparency, fairness, reliability and safety, privacy and security, and inclusiveness goals).
[20] *See, e.g.*, Jessica Newman, Univ. of Cal. Berkeley Center for Long-Term Cybersecurity, A Taxonomy of Trustworthiness for Artificial Intelligence White Paper (Jan. 2023), Univ. of Cal. Berkeley Center for Long-Term Cybersecurity, https://cltc.berkeley.edu/wp-content/uploads/2023/01/Taxonomy_of_AI_Trustworthiness.pdf (mapping 150 properties of trustworthiness, building on NIST AI Risk Management Framework); Thilo Hagendorff, The Ethics of AI Ethics: An Evaluation of Guidelines, Minds & Machines 30, 99–120 (2020), https://doi.org/10.1007/s11023-020-09517-8; Jeannette M. Wing, Trustworthy AI, Communications of the ACM, Vol. 64 No. 10 (Oct. 2021), https://cacm.acm.org/magazines/2021/10/255716-trustworthy-ai/fulltext.
[21] *See generally,* Luciano Floridi, and Josh Cowls, A Unified Framework of Five Principles for AI in Society, Harvard Data Science Review, Issue1.1 (July 01, 2019), https://doi.org/10.1162/99608f92.8cd550d1 (synthesizing ethical AI codes); Algorithm Watch, The AI Ethics Guidelines Global Inventory (2022), https://inventory.algorithmwatch.org (last visited Feb. 1, 2023) (listing 165 sets of ethical AI guidelines).
[22] *See, e.g.*, Institute of Electrical and Electronics Engineers (IEEE), IEEE Global Initiative on Ethics of Autonomous & Intelligent Systems (Feb. 2022), http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf; IEEE, IEEE P7014: Emulated Empathy in Autonomous and Intelligent Systems Working Group, https://sagroups.ieee.org/7014 (last visited Feb. 1, 2023). *C.f.* Daniel Schiff et al., , IEEE 7010: A New Standard for Assessing the Well-Being Implications of Artificial Intelligence, IEEE Int'l Conf. on Sys., Man & Cybernetics 1 (2020). There also efforts to harmonize and compare tools for trustworthy AI. *See, e.g.,* OECD, OECD Tools for Trustworthy AI: A Framework to Compare Implementation Tools for Trustworthy AI Systems, OECD Digital Economy Papers No. 312 (June 2021), https://www.oecd-ilibrary.org/docserver/008232ec-en.pdf?expires=1674495915&id=id&accname=guest&checksum=F5D10D29FCE205F3F32F409A679571FE.
[23] *See, e.g.*, European Commission, High-Level Expert Group on Artificial Intelligence (AI HLEG), Ethics Guidelines for Trustworthy AI (Apr. 8, 2019), https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

particular scrutiny is warranted because of their unique features and fast-growing importance in American life and commerce. As NIST notes, these systems are

> "trained on data that can change over time, sometimes significantly and unexpectedly, affecting system functionality and trustworthiness in ways that are hard to understand. AI systems and the contexts in which they are deployed are frequently complex, making it difficult to detect and respond to failures when they occur. AI systems are inherently socio-technical in nature, meaning they are influenced by societal dynamics and human behavior. AI risks – and benefits – can emerge from the interplay of technical aspects combined with societal factors related to how a system is used, its interactions with other AI systems, who operates it, and the social context in which it is deployed."[24]

The objective of this engagement is to solicit input from stakeholders in the policy, legal, business, academic, technical, and advocacy arenas on how to develop a productive AI accountability ecosystem. Specifically, NTIA hopes to identify the state of play, gaps, and barriers to creating adequate accountability for AI systems, any trustworthy AI goals that might not be amenable to requirements or standards, how supposed accountability measures might mask or minimize AI risks, the value of accountability mechanisms to compliance efforts, and ways governmental and non-governmental actions might support and enforce AI accountability practices.

[24] AI Risk Mgmt. Framework 1.0, *supra* note 11 at 1.

This Request for Comment uses the terms AI, algorithmic, and automated decision systems without specifying any particular technical tool or process. It incorporates NIST's definition of an "AI system," as "an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments."[25] This Request's scope and use of the term "AI" also encompasses the broader set of technologies covered by the Blueprint: "automated systems" with "the potential to meaningfully impact the American public's rights, opportunities, or access to critical resources or services."[26]

Accountability for Trustworthy AI

1. Growing Regulatory Interest in AI Accountability Mechanisms

Governments, companies, and civil society organizations are developing AI governance tools to mitigate the risks of autonomous systems to individuals and communities. Among these are accountability mechanisms to show that AI systems are trustworthy, which can help foster responsible development and deployment of algorithmic systems, while at the same time giving affected parties (including customers, investors, affected individuals and communities, and regulators) confidence that the technologies are in fact worthy of trust.[27] Governments around the world, and within the United States, are beginning to require accountability mechanisms including audits and assessments of AI systems, depending upon their use case and risk level. For example, there are relevant provisions in the European Union's Digital Services Act

[25] *Id*.
[26] Blueprint for AIBoR, *supra* note 10 at 8.
[27] *See, e.g.,* Michael Kearns and Aaron Roth, Ethical Algorithm Design Should Guide Technology Regulation, Brookings (Jan. 13, 2020), https://www.brookings. edu/research/ethical-algorithm-design-should-guide-technology-regulation (noting that "more systematic, ongoing, and legal ways of auditing algorithms are needed").

requiring audits of very large online platforms' systems,[28] the draft EU Artificial Intelligence Act requiring conformity assessments of certain high-risk AI tools before deployment,[29] and New York City Law 144 requiring bias audits of certain automated hiring tools used within its jurisdiction.[30] Several bills introduced in the U.S. Congress include algorithmic impact assessment or audit provisions.[31] In the data and consumer protection space, policies focus on design features of automated systems by requiring in the case of privacy-by-design,[32] or

---

[28] European Union, Amendments Adopted by the European Parliament on 20 January 2022 on the Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, OJ C 336, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022AP0014 (Article 28 provides that "[v]ery large online platforms shall ensure auditors have access to all relevant data necessary to perform the audit properly." Further, auditors must be "recognised and vetted by the Commission and … [must be] legally and financially independent from, and do not have conflicts of interest with" the audited platforms.).

[29] European Union, Proposal for a Regulation of the European Parliament and Of The Council Laying Down Harmonised Rules On Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, 2021/0106(COD), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206. *See also* European Parliament Special Committee on Artificial Intelligence in a Digital Age*,* Report on Artificial Intelligence in a Digital Age, A9-0088/2022, https://www.europarl.europa.eu/doceo/document/A-9-2022-0088_EN.html (setting forth European Parliament positions on AI development and governance).

[30] The New York City Council, Automated Employment Decision Tools, Int 1894-2020 (effective Apr. 2023), https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9&Options=ID%7CText%7C&Search=. A similar law has been proposed in New Jersey. Bill A4909 (Sess. 2022-2023), https://legiscan.com/NJ/text/A4909/2022. *See also*, Colorado SB 21-169, Protecting Consumers from Unfair Discrimination in Insurance Practices (2021) (requiring insurers to bias test big data systems, including algorithms and predictive models, and to demonstrate testing methods and nondiscriminatory results to the Colorado Division of Insurance); State of Connecticut Insurance Dept., Notice to All Entities and Persons Licensed by the Connecticut Insurance Department Concerning the Usage of Big Data and Avoidance of Discriminatory Practices (April 20, 2022) (expressing potential regulatory concerns with "[h]ow Big Data algorithms, predictive models, and various processes are inventoried, risk assessed/ranked, risk managed, validated for technical quality, and governed throughout their life cycle to achieve the mandatory compliance" with non-discrimination laws and reminding insurers to submit annual data certifications), https://portal.ct.gov/-/media/CID/1_Notices/Technologie-and-Big-Data-Use-Notice.pdf.

[31] *See, e.g*., American Data Privacy and Protection Act, H.R.8152, 117th Cong. § 207(c) (2022) (proposing to require large data holders using covered algorithms posing consequential risk of harm to individuals or groups to conduct risk assessment and report on risk mitigation measures); Algorithmic Accountability Act of 2022, H.R. 6580, 117th Cong. (2022) (would require covered entities to produce impact assessments for the Federal Trade Commission).

[32] *See, e.g.,* Council 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. 25 (implementing data protection by design principles).

prohibiting in the case of "dark patterns," certain design choices to secure data and consumer protection.[33] Governments are mandating accountability measures for government-deployed AI systems.[34] Related tools are also emerging in the private sector from non-profit entities such as the Responsible AI Institute (providing system certifications)[35] to startups and well-established companies, such as Microsoft's Responsible AI Standard[36] and Datasheets for Datasets,[37] the Rolls-Royce Alethia Framework,[38] Google's Model Card Toolkit,[39] and many others.

Federal regulators have been addressing AI system risk management in certain sectors for more than a decade. For example, the Federal Reserve in 2011 issued SR-11-7 Guidance on Algorithmic Model Risk Management, noting that reducing risks requires "critical analysis by objective, informed parties that can identify model limitations and produce appropriate changes" and, relatedly, the production of testing, validation, and associated records for examination by independent parties.[40] As financial agencies continue to explore AI accountability mechanisms in

---

[33] *See, e.g.,* Cal. Civ. Code §1798.140, subd. (l), (h) (effective Jan. 1, 2023) (regulating the use of a "dark pattern" defined as a "user interface designed or manipulated with the substantial effect of subverting or impairing user autonomy, decision-making, or choice, as further defined by regulation" and noting that "agreement obtained through use of dark patterns does not constitute consent.").

[34] *See, e.g.,* Treasury Board of Canada Secretariat, Algorithmic Impact Assessment Tool, Government of Canada (modified April 19, 2022), https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html; Treasury Board of Canada Secretariat, Directive on Automated Decision-Making, Government of Canada (modified April 1, 2021), https://www.tbssct.canada.ca/pol/doc-eng.aspx?id=32592.

[35] Responsible Artificial Intelligence Institute, https://www.responsible.ai/ (last visited Apr. 2, 2023).

[36] Microsoft, Microsoft Responsible AI Standard, v2 General Requirements (June 2022), https://blogs.microsoft.com/wp-content/uploads/prod/sites/5/2022/06/Microsoft-Responsible-AI-Standard-v2-General-Requirements-3.pdf.

[37] Microsoft, Aether Data Documentation Template (Draft 08/25/2022), https://www.microsoft.com/en-us/research/uploads/prod/2022/07/aether-datadoc-082522.pdf. *See also* Timnit Gebru et. Al., Datasheets for Datasets, Communications of the ACM,Vol. 64, No. 12, 86– 92 (Dec. 2021).

[38] Rolls Royce, Aletheia Framework, https://www.rolls-royce.com/innovation/the-aletheia-framework.aspx (last visited Mar. 3, 2023).

[39] GitHub, Tensorflow/model-card-toolkit, https://github.com/tensorflow/model-card-toolkit (last visited Jan. 30, 2023) ("A toolkit that streamlines and automates the generation of model cards").

[40] Board of Governors of the Federal Reserve System, Supervisory Guidance on Model Risk Management, Federal Reserve SR Letter 11-7 (Apr. 4, 2011), https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm.

their areas,[41] other federal agencies such as the Equal Employment Opportunities Commission have begun to do the same.[42] Moreover, state regulators are considering compulsory AI accountability mechanisms.[43]

2. AI Audits and Assessments

AI systems are being used in human resources and employment, finance, health care, education, housing, transportation, law enforcement and security, and many other contexts that significantly impact people's lives. The appropriate goal and method to advance AI accountability will likely depend on the risk level, sector, use case, and legal or regulatory requirements associated with the system under examination. Assessments and audits are among the most common mechanisms to provide assurance about AI system characteristics. Guidance, academic, and regulatory documents use the terms "assessments" (including risk, impact, and conformity) and "audits" in various ways and without standard definition.[44] Often in these references, "assessment" refers to an entity's internal review of an AI system to identify risks or outcomes. An "audit" often refers to an external review of an AI system at a point in time to

[41] Department of Treasury, Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, Bureau of Consumer Financial Protection, and National Credit Union Administration, Request for Information and Comment on Financial Institutions' Use of Artificial Intelligence, Including Machine Learning, 86 Fed. Reg. 16837 (Mar. 31, 2021).

[42] *See* U.S. Equal Employment Opportunity Commission, The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees (May 12, 2022) (issuing technical guidance on algorithmic employment decisions in connection with the Americans with Disabilities Act), https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence.

[43] *See, e.g.,* Colorado Department of Regulatory Agencies Division of Insurance, Draft Proposed New Regulation: Governance and Risk Management Framework Requirements for Life Insurance Carriers' Use of External Consumer Data and Information Sources, Algorithms and Predictive models (Feb. 1, 2023), https://protect-us.mimecast.com/s/V0LqCVOVw1Hl6g5xSNSwGG?domain=lnks.gd.

[44] *See, e.g.,* Louis An Yeung, Guidance for the Development of AI Risk & Impact Assessments*,* Center for Long-Term Cybersecurity (July 2021), at 5, https://cltc.berkeley.edu/wp-content/uploads/2021/08/AI_Risk_Impact_Assessments.pdf (surveying definitions and concluding that AI risk and impact assessments "may be used interchangeably.").

assess performance against accepted benchmarks. Assessments and audits may both be conducted on a continuous basis, and may be conducted either by internal or external reviewers.

Common areas of focus for AI audits and assessments include harmful bias and discrimination, effectiveness and validity, data protection and privacy, and transparency and explainability (how understandable AI system predictions or decisions are to humans). For information services like social media, large language and other generative AI models, and search, audits and assessments may also cover harms related to the distortion of communications through misinformation, disinformation, deep fakes, privacy invasions, and other content-related phenomena.[45]

Audits may be conducted internally or by independent third parties.[46] An internal audit may be performed by the team that developed the technology or by a separate team within the same entity. Independent audits may range from "black box" adversarial audits conducted without the help of the audited entity[47] to "white box" cooperative audits conducted with substantial access to the relevant models and processes.[48] Audits may be made public or given limited circulation,

[45] Jack Bandy, Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits, Proceedings of the ACM on Human-Computer Interaction, Vol.5. No. 74, 1–34 (April 2021), https://doi.org/10.1145/3449148 (identifying discrimination and distortion as the most commonly audited-for outputs of algorithm systems).
[46] Responsible Artificial Intelligence Institute, Responsible AI Certification Program - White Paper (Oct. 2022), https://assets.ctfassets.net/rz1q59puyoaw/5pyXogKSKNUKRkqOP4hRfy/5c5b525d0a77a1017643dcb6b5124634/RAII_Certification_Guidebook.pdf.
[47] Danae Metaxa et al., Auditing Algorithms: Understanding Algorithmic Systems from the Outside In, ACL Digital Library (Nov. 25, 2021), https://dl.acm.org/doi/10.1561/1100000083.
[48] *See, e.g.,* Christo Wilson, et. al., Building and Auditing Fair Algorithms: A Case Study in Candidate Screening, FAccT '21 (March 1–10, 2021), https://evijit.github.io/docs/pymetrics_audit_FAccT.pdf.

for example to regulators.[49] They may be conducted by professional experts or undertaken by impacted lay people.[50]

While some audits and assessments may be limited to technical aspects of a particular model, it is widely understood that AI models are part of larger systems, and these systems are embedded in socio-technical contexts. How models are implemented in practice could depend on model interactions, employee training and recruitment, enterprise governance, stakeholder mapping and engagement,[51] human agency, and many other factors.[52] The most useful audits and assessments of these systems, therefore, should extend beyond the technical to broader questions about governance and purpose. These might include whether the people affected by AI systems are meaningfully consulted in their design[53] and whether the choice to use the technology in the first place was well-considered.[54]

---

[49] *See, e.g.*, Council of the District of Columbia, Stop Discrimination by Algorithms Act of 2021, B24-558, https://oag.dc.gov/sites/default/files/2021-12/DC-Bill-SDAA-FINAL-to-file-.pdf (proposing law that would require audits of certain algorithmic systems to be shared with the Attorney General of the District of Columbia).

[50] *See, e.g.,* Michelle S. Lam et al., End-User Audits: A System Empowering Communities to Lead Large-Scale Investigations of Harmful Algorithmic Behavior, Proceedings of the ACM Human-Computer Interaction, Vol. 6, Issue CSCW2, Article 512, 1-32 (November 2022), https://doi.org/10.1145/3555625 (describing an "end-user audit" deployed in the content moderation setting to audit Perspective API toxicity predictions).

[51] *See, e.g.,* Alan Turing Institute*,* Human Rights, Democracy, and the Rule of Law Assurance Framework for AI Systems: A Proposal Prepared for the Council of Europe's Ad hoc Committee on Artificial Intelligence, 211-223 (2021), https://rm.coe.int/huderaf-coe-final-1-2752-6741-5300-v-1/1680a3f688 (exemplifying what stakeholder mapping might entail).

[52] *See, e.g.,* Inioluwa Deborah Raji et al., Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing, FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33-44, 37 (January 2020), https://doi.org/10.1145/3351095.3372873; Inioluwa Deborah Raji et al., Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance, AIES '22: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society 560, 566 (June 9, 2022), https://dl.acm.org/doi/pdf/10.1145/3514094.3534181.

[53] Adriano Koshiyama et al., Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms (Feb. 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3778998.

[54] *See, e.g.*, Alene Rhea et al., Resume Format, LinkedIn URLs and Other Unexpected Influences on AI Personality Prediction in Hiring: Results of an Audit, Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22), Association for Computing Machinery, 572–587 (July 2022),

Some accountability mechanisms may use legal standards as a baseline. For example, standards for employment discrimination on the basis of sex, religion, race, color, disability, or national origin may serve as benchmarks for AI audits,[55] as well as for legal compliance actions.[56] Civil society groups are developing additional operational guidance based on such standards.[57] Some firms and startups are beginning to offer testing of AI models on a technical level for bias and/or disparate impact. It should be recognized that for some features of trustworthy AI, consensus standards may be difficult or impossible to create.

3. Policy Considerations for the AI Accountability Ecosystem

https://doi.org/10.1145/3514094.3534189 (finding that personality tests used in automated hiring decisions cannot be considered valid); Sarah Bird, Responsible AI Investments and Safeguards for Facial Recognition, Microsoft Azure, (June 21, 2022), https://azure.microsoft.com/en-us/blog/responsible-ai-investments-and-safeguards-for-facial-recognition (announcing phase-out of emotion recognition from Azure Face API facial recognition services because of lack of evidence of effectiveness).

[55] *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 (auditing the claims of an automated hiring tool that it satisfied Title VII of the Civil Rights Act's four-fifths rule). *C.f.* Pauline Kim, Data-Driven Discrimination at Work*,* 58 Wm. & Mary L. Rev. 857 (2017) (addressing limitations of Title VII liability provisions as an adequate means to prevent classification bias in hiring); U.S. Equal Employment Opportunity Commission, Navigating Employment Discrimination in AI and Automated Systems: A New Civil Rights Frontier, Meetings of the Commission, Testimony of Manish Raghavan, (Jan. 31, 2023), https://www.eeoc.gov/meetings/meeting-january-31-2023-navigating-employment-discrimination-ai-and-automated-systems-new/raghavan (highlighting data-related and other challenges of auditing AI systems used in hiring according to the four-fifths rule).

[56] *See, e.g*., U.S. Equal Employment Opportunity Commission, The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence to Assess Job Applicants and Employees (May 12, 2022) (issuing technical guidance on algorithmic employment decisions in connection with the Americans with Disabilities Act), https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence; U.S. Department of Justice, Justice Department Files Statement of Interest in Fair Housing Act Case Alleging Unlawful Algorithm-Based Tenant Screening Practices, Press Release (Jan. 9, 2023), https://www.justice.gov/opa/pr/justice-department-files-statement-interest-fair-housing-act-case-alleging-unlawful-algorithm.

[57] *See, e.g.,* Matt Scherer and Ridhi Shetty, Civil Rights Standards for 21st Century Employment Selection Procedures, Center for Democracy and Technology, (Dec. 2022), https://cdt.org/insights/civil-rights-standards-for-21st-century-employment-selection-procedures (guidance on pre-deployment and post-deployment audits and assessments of algorithmic tools in the employment context to detect and mitigate adverse impacts on protected classes).

Among the challenges facing policymakers in the AI accountability space are tradeoffs among trustworthy AI goals, barriers to implementing accountability mechanisms, complex AI lifecycle and value chains, and difficulties with standardization and measurement.

Accountability ecosystems that might serve as models for AI systems range from financial assurance, where there are relatively uniform financial auditing practices,[58] to environmental, social, and governance (ESG) assurance, where standards are quite diverse.[59] Considering the range of trustworthy AI system goals and deployment contexts, it is likely that at least in the near term, AI accountability mechanisms will be heterogeneous. Commentators have raised concerns about the validity of certain accountability measures. Some audits and assessments, for example, may be scoped too narrowly, creating a "false sense" of assurance.[60] Given this risk, it is imperative that those performing AI accountability tasks are sufficiently qualified to provide credible evidence that systems are trustworthy.[61]

There may be other barriers to providing adequate and meaningful accountability. Some mechanisms may require datasets built with sensitive data that puts privacy or security at risk, raising questions about trade-offs among different values. In addition, there may be insufficient

---

[58] *See generally,* Financial Accounting Standards Board, Generally Accepted Accounting Principles, http://asc.fasb.org/home.

[59] *See generally,* Elizabeth Pollman, "Corporate Social Responsibility, ESG, and Compliance" in Benjamin Van Rooij and D. Daniel Sokol (Eds.) The Cambridge Handbook of Compliance (2021) ("Companies have flexibility to create their own structures for internal governance, their own channels for stakeholder engagement, their own selection of third-party guidelines or standards, and in many jurisdictions, their own level of disclosure.").

[60] *See, e.g.,* Brandie Nonnecke and Philip Dawson, Human Rights Implications of Algorithmic Impact Assessments: Priority Considerations to Guide Effective Development and Use, Harvard Kennedy School - Carr Center for Human Rights Policy, Carr Center Discussion Paper (Oct. 21, 2021), https://carrcenter.hks.harvard.edu/files/cchr/files/nonnecke_and_dawson_human_rights_implications.pdf.

[61] *See, e.g.,* Sasha Costanza-Chock et al., Who Audits the Auditors? Recommendations from a Field Scan of the Algorithmic Auditing Ecosystem, FaccT'22: Proceedings of the 2022 Association for Computing Machinery Conference on Fairness, Accountability, and Transparency, 1571-1583 (June 21–24, 2022), https://doi.org/10.1145/3531146.3533213.

access to the subject system or its data, insufficient qualified personnel to audit systems, and/or inadequate audit or assessment standards to benchmark the work.[62]

Timing is another complication for AI accountability, and especially for providing assurance of AI systems. The point in an AI system lifecycle at which an audit or assessment is conducted, for example, will impact what questions it answers, how much accountability it provides, and to whom that accountability is offered. The General Services Administration has depicted an AI lifecycle that starts with pre-design (*e.g.*, problem specification, data identification, use case selection), progresses through design and development (*e.g.*, model selection, training, and testing), and then continues through deployment.[63] Other federal agencies use substantially similar lifecycle schema.[64] Throughout this lifecycle, dynamic interactions with data and iterative learning create many moments for evaluation of specific models and the AI system as a whole.[65]

The AI value chain, including data sources, AI tools, and the relationships among developers and customers, can also be complicated and impact accountability. Sometimes a developer will train an AI tool on data provided by a customer, or the customer may in turn use the tool in ways the developer did not foresee or intend. Data quality is an especially important variable to

---

[62] Centre for Data Ethics and Innovation, Industry Temperature Check: Barriers and Enablers to AI Assurance (Dec. 2022), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1122115/Industry_Temperature_Check_-_Barriers_and_Enablers_to_AI_Assurance.pdf.

[63] For other lifecycle models, *see* International Organization for Standardization, Information Technology—Artificial Intelligence—AI System Life Cycle Processes (ISO/IEC DIS 5338), Edition1, https://www.iso.org/standard/81118.html (under development as of Oct. 22, 2022).

[64] *See, e.g.,* U.S. Department of Energy, DOE AI Risk Management Playbook, https://www.energy.gov/ai/doe-ai-risk-management-playbook-airmp (last visited Jan 30, 2023) (identifying AI lifecycle stages as (0) problem identification, (1) supply chain, (2) data acquisition, (3) model development, (4) model deployment, and (5) model performance).

[65] *See generally*, Norberto Andrade et al., Artificial Intelligence Act: A Policy Prototyping Experiment - Operationalizing the Requirements for AI Systems – Part I, 24-33 (Nov. 2022) https://openloop.org/reports/2022/11/Artificial_Intelligence_Act_A_Policy_Prototyping_Experiment_Operationalizing_Reqs_Part1.pdf (providing examples of interactions between data and algorithmic outputs along the AI lifecycle and value chain).

examine in AI accountability.[66] A developer training an AI tool on a customer's data may not be able to tell how that data was collected or organized, making it difficult for the developer to assure the AI system. Alternatively, the customer may use the tool in ways the developer did not foresee or intend, creating risks for the developer wanting to manage downstream use of the tool. When responsibility along this chain of AI system development and deployment is fractured, auditors must decide whose data and which relevant models to analyze, whose decisions to examine, how nested actions fit together, and what is within the audit's frame.

Public and private bodies are working to develop metrics or benchmarks for trustworthy AI where needed.[67] Standards-setting bodies such as IEEE[68] and ISO,[69] as well as research organizations focusing on measurements and standards, notably NIST,[70] are devising technical standards that can improve AI governance and risk management and support AI accountability. These include standards for general technology process management (*e.g.,* risk management),

---

[66] *See, e.g.,* European Union Agency for Fundamental Rights, Data Quality and Artificial Intelligence – Mitigating Bias and Error to Protect Fundamental Rights (June 7, 2019), https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-data-quality-and-ai_en.pdf (noting the importance for managing downstream risk of high-quality data inputs, including completeness, accuracy, consistency, timeliness, duplication, validity, availability, and whether the data are fit for the purpose).

[67] *See, e.g.,* Centre for Data Ethics and Innovation, The Roadmap to an Effective AI Assurance Ecosystem – extended version (Dec 8, 2021), https://www.gov.uk/government/publications/the-roadmap-to-an-effective-ai-assurance-ecosystem/the-roadmap-to-an-effective-ai-assurance-ecosystem-extended-version; Digital Regulation Cooperation Forum, Auditing algorithms: The Existing Landscape, Role of Regulators and Future Outlook (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.

[68] *See e.g.,* Institute of Electrical and Electronics Engineers Standards Association, CertifAIEd*,* https://engagestandards.ieee.org/ieeecertifaied.html (last visited Jan 31, 2023) (a certification program for assessing ethics of Autonomous Intelligent Systems).

[69] *See, e.g.*, International Organization for Standardization, Information Technology—Artificial intelligence—Transparency Taxonomy of AI Systems (ISO/IEC AWI 12792), Edition 1, https://www.iso.org/standard/84111.html (under development as of Jan. 30, 2023).

[70] *See, e.g.,* NIST, AI Standards: Federal Engagement, https://www.nist.gov/artificial-intelligence/ai-standards-federal-engagement (last visited Jan 31, 2023) (committing to standards work related to accuracy, explainability and interpretability, privacy, reliability, robustness, safety, security resilience, and anti-bias so as to "help the United States to speed the pace of reliable, robust, and trustworthy AI technology development.").

standards applicable across technologies and applications (*e.g.*, transparency and anti-bias), and standards for particular technologies *(e.g.*, emotion detection and facial recognition). For some trustworthy AI goals, it will be difficult to harmonize standards across jurisdictions or within a standard-setting body, particularly if the goal involves contested moral and ethical judgements. In some contexts, *not* deploying AI systems at all will be the means to achieve the stated goals.

To address these barriers and complexities, commentators have suggested that policymakers and others can foster AI accountability by: mandating impact assessments[71] and audits,[72] defining "independence" for third-party audits,[73] setting procurement standards,[74] incentivizing effective audits and assessments through bounties, prizes, and subsidies,[75] creating access to data

[71] *See, e.g.,* Andrew D. Selbst, An Institutional View of Algorithmic Impact Assessments, 35 Harv. J.L. & Tech. 117 (2021).

[72] *See, e.g.,* Alex Engler, How the Biden Administration Should Tackle AI Oversight, Brookings (Dec. 10, 2020), https://www.brookings.edu/research/how-the-biden-administration-should-tackle-ai-oversight (advocating government audits of "highly impactful, large-scale AI systems"); Danielle Keats Citron and Frank Pasquale, The Scored Society: Due Process for Automated Predictions, 89 Wash U. L. Rev.1, 20-22 (2014) (advocating audit requirements for algorithmic systems used in employment, insurance, and health care contexts).

[73] *See, e.g.,* Ifeoma Ajunwa, An Auditing Imperative for Automated Hiring Systems, 34 Harv. J. L. & Tech 621, 668-670 (2021).

[74] *See, e.g.,* Deirdre K. Mulligan and Kenneth A. Bamberger, Procurement as Policy: Administrative Process for Machine Learning, 34 Berkeley Tech. L. J. 773, 841-44 (2019) (discussing public procurement processes); Jennifer Cobbe et al., Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems, Proceedings of the 2021 Association for Computing Machinery Conference on Fairness, Accountability, and Transparency, 598-609, 604 (March 2021), https://dl.acm.org/doi/10.1145/3442188.3445921 (discussing relevance of procurement records to accountability relationships).

[75] *See, e.g.,* Miles Brundage et al., Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims, arXiv, 16-17, (April 20, 2020) https://arxiv.org/abs/2004.07213 (proposing the expanded use of bounties to help detect safety, bias, privacy, and other problems with AI systems); *see also* Rumman Chowdhury and Jutta Williams, Introducing Twitter's First Algorithmic Bias Bounty Challenge, Twitter Engineering (Jul. 30, 2021), https://blog.twitter.com/engineering/en_us/topics/insights/2021/algorithmic-bias-bounty-challenge.

necessary for AI audits and assessments,[76] creating consensus standards for AI assurance,[77] providing auditor certifications,[78] and making test data available for use.[79] We particularly seek input on these policy proposals and mechanisms.

Instructions for Commenters:

Through this Request for Comment, we hope to gather information on the following questions. These are not exhaustive, and commenters are invited to provide input on relevant questions not asked below. Commenters are not required to respond to all questions. When responding to one or more of the questions below, please note in the text of your response the number of the question to which you are responding. Commenters should include a page number on each page of their submissions. Commenters are welcome to provide specific actionable proposals, rationales, and relevant facts.

Please do not include in your comments information of a confidential nature, such as sensitive personal information or proprietary information. All comments received are a part of

---

[76] *See, e.g.*, Sonia González-Bailón, & Yphtach Lelkes, Do Social Media Undermine Social Cohesion? A Critical Review, Social Issues and Policy Review, Vol. 17, Issue 1, 1-180, 21 (2022), https://doi.org/10.1111/sipr.12091 (arguing that for investigations of social media algorithms, "[p]olicy makers should consider sponsoring academic-industry partnerships allowing researchers to access this research and the data generated in the process to produce evidence of public value while securing privacy").

[77] *See, e.g.*, Jakob Mökander and Maria Axente. Ethics-Based Auditing of Automated Decision-Making Systems: Intervention Points and Policy Implication*s*, AI & Society, 28, 153-171 (Oct. 2021), https://doi.org/10.1007/s00146-021-01286-x.

[78] *See, e.g.,* United Nations Educational, Scientific and Cultural Organization (UNESCO), Recommendation on the Ethics of Artificial Intelligence (Nov. 23, 2021) at 27, https://unesdoc.unesco.org/ark:/48223/pf0000380455 ("Member States are encouraged to …consider forms of soft governance such as a certification mechanism for AI systems and the mutual recognition of their certification, according to the sensitivity of the application domain and expected impact on human rights, the environment and ecosystems, and other ethical considerations …[including] different levels of audit of systems, data, and adherence to ethical guidelines and to procedural requirements in view of ethical aspects.").

[79] *See, e.g.,* National Artificial Intelligence Research Resource Task Force, Strengthening and Democratizing the U.S. Artificial Intelligence Innovation Ecosystem: An Implementation Plan for a National Artificial Intelligence Research Resource, 32-36 (Jan. 2023), https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf (proposing the federal curation of datasets for use in training and testing AI systems).

the public record and will generally be posted to Regulations.gov without change. All personal identifying information (*e.g.*, name, address) voluntarily submitted by the commenter may be publicly accessible.

Questions:

AI Accountability Objectives

1. What is the purpose of AI accountability mechanisms such as certifications, audits, and assessments? Responses could address the following:
    a. What kinds of topics should AI accountability mechanisms cover? How should they be scoped?
    b. What are assessments or internal audits most useful for? What are external assessments or audits most useful for?
    c. An audit or assessment may be used to verify a claim, verify compliance with legal standards, or assure compliance with non-binding trustworthy AI goals. Do these differences impact how audits or assessments are structured, credentialed, or communicated?
    d. 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?
    e. 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?

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?

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?
    a. The AI system does not substantially contribute to harmful discrimination against people.
    b. The AI system does not substantially contribute to harmful misinformation, disinformation, and other forms of distortion and content-related harms.
    c. The AI system protects privacy.
    d. The AI system is legal, safe, and effective.
    e. There has been adequate transparency and explanation to affected people about the uses, capabilities, and limitations of the AI system.
    f. There are adequate human alternatives, consideration, and fallbacks in place throughout the AI system lifecycle.
    g. There has been adequate consultation with, and there are adequate means of contestation and redress for, individuals affected by AI system outputs.
    h. There is adequate management within the entity deploying the AI system such that there are clear lines of responsibility and appropriate skillsets.

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?

5. Given the likely integration of generative AI tools such as large language models (*e.g.*, ChatGPT) or other general-purpose AI or foundational models into downstream products, how can AI accountability mechanisms inform people about how such tools are operating and/or whether the tools comply with standards for trustworthy AI?[80]

6. The application of accountability measures (whether voluntary or regulatory) is more straightforward for some trustworthy AI goals than for others. With respect to which trustworthy AI goals are there existing requirements or standards? Are there any trustworthy AI goals that are not amenable to requirements or standards? How should accountability policies, whether governmental or non-governmental, treat these differences?

7. 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?

[80] *See, e.g.,* Jakob Mökander et al., Auditing Large Language Models: A Three-layered Approach (preprint 2003), ArXiv, https://doi.org/10.48550/ARXIV.2302.08500.

8. What are the best definitions of and relationships between AI accountability, assurance, assessments, audits, and other relevant terms?

Existing Resources and Models

9. 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?

10. What are the best definitions of terms frequently used in accountability policies, such as fair, safe, effective, transparent, and trustworthy? Where can terms have the same meanings across sectors and jurisdictions? Where do terms necessarily have different meanings depending on the jurisdiction, sector, or use case?

11. What lessons can be learned from accountability processes and policies in cybersecurity, privacy, finance, or other areas?[81]

[81] *See, e.g.,* Megan Gray, Understanding and Improving Privacy 'Audits' Under FTC Orders (April 18, 2018), at 4-8, http://dx.doi.org/10.2139/ssrn.3165143 (critiquing the implementation of third-party privacy audit mandates). For an example of more recent provisions for privacy audits, *see United States v. Epic Games*, Stipulated Order for Permanent Injunction, Civ. No. 5:22-cv-00518-BO (E.D.N.C. Dec. 19, 2022), 22-25 (requiring assessments by independent third-party auditors in a children's privacy settlement), https://www.ftc.gov/system/files/ftc_gov/pdf/2223087EpicGamesSettlement.pdf.

12. What aspects of the United States and global financial assurance systems provide useful and achievable models for AI accountability?

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

14. Which non-U.S. or U.S. (federal, state, or local) laws and regulations already requiring an AI audit, assessment, or other accountability mechanism are most useful and why? Which are least useful and why?

Accountability Subjects

15. The AI value or supply chain is complex, often involving open source and proprietary products and downstream applications that are quite different from what AI system developers may initially have contemplated. Moreover, training data for AI systems may be acquired from multiple sources, including from the customer using the technology. Problems in AI systems may arise downstream at the deployment or customization stage or upstream during model development and data training.
    a. Where in the value chain should accountability efforts focus?
    b. How can accountability efforts at different points in the value chain best be coordinated and communicated?
    c. How should vendors work with customers to perform AI audits and/or assessments? What is the role of audits or assessments in the commercial and/or

public procurement process? Are there specific practices that would facilitate credible audits (*e.g.*, liability waivers)?

d. 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?

16. The lifecycle of any given AI system or component also presents distinct junctures for assessment, audit, and other measures. For example, in the case of bias, it has been shown that "[b]ias is prevalent in the assumptions about which data should be used, what AI models should be developed, where the AI system should be placed — or if AI is required at all."[82] How should AI accountability mechanisms consider the AI lifecycle? Responses could address the following:

    a. 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?[83] 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?

[82] Reva Schwartz et al., Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*,* NIST Special Publication 1270, at 6, https://doi.org/10.6028/NIST.SP.1270.

[83] *See generally,* Inioluwa Deborah Raji and Joy Buolamwini, Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products, AIES 2019 - Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 429–435 (2019), https://doi.org/10.1145/3306618.3314244 (discussing scoping questions).

b. How should AI audits or assessments be timed? At what stage of design, development, and deployment should they take place to provide meaningful accountability?
c. How often should audits or assessments be conducted, and what are the factors that should inform this decision? How can entities operationalize the notion of continuous auditing and communicate the results?
d. What specific language should be incorporated into governmental or non-governmental policies to secure the appropriate timing of audits or assessments?

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?

18. Should AI systems be released with quality assurance certifications, especially if they are higher risk?

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

Accountability Inputs and Transparency

20. What sorts of records (*e.g.*, logs, versions, model selection, data selection) and other documentation should developers and deployers of AI systems keep in order to support AI accountability?[84] How long should this documentation be retained? Are there design principles (including technical design) for AI systems that would foster accountability-by-design?

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?

22. How should the accountability process address data quality and data voids of different kinds? For example, in the context of automated employment decision tools, there may be no historical data available for assessing the performance of a newly deployed, custom-built tool. For a tool deployed by other firms, there may be data a vendor has access to, but the audited firm itself lacks. In some cases, the vendor itself may have intentionally limited its own data collection and access for privacy and security purposes. How should AI accountability requirements or practices deal with these data issues? What should be the roles of government, civil society, and academia in

[84] *See, e.g.,* Miles Brundage et al., Toward Trustworthy AI Development: Mechanisms for Supporting Verifiable Claims at 24-25 (2020), http://www.towardtrustworthyai.com/ (last visited Jan 30, 2023) (discussing audit trail components). *See also* AI Risk Mgmt. Framework 1.0, *supra* note 11 at 15 (noting that transparent AI informs individuals about system characteristics and functions ranging from "design decisions and training data to model training, the structure of the model, its intended use cases, and how and when deployment, post-deployment, or end user decisions were made and by whom"); *id*. at 16 (defining related terms: "Explainability refers to a representation of the mechanisms underlying AI systems' operation, whereas *interpretability* refers to the meaning of AI systems' output in the context of their designed functional purposes").

providing useful data sets (synthetic or otherwise) to fill gaps and create equitable access to data?

23. How should AI accountability "products" (*e.g.*, audit results) be communicated to different stakeholders? Should there be standardized reporting within a sector and/or across sectors? How should the translational work of communicating AI accountability results to affected people and communities be done and supported?

Barriers to Effective Accountability

24. What are the most significant barriers to effective AI accountability in the private sector, including barriers to independent AI audits, whether cooperative or adversarial? What are the best strategies and interventions to overcome these barriers?

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

26. Is the lack of a federal law focused on AI systems a barrier to effective AI accountability?

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?

28. What do AI audits and assessments cost? Which entities should be expected to bear these costs? What are the possible consequences of AI accountability requirements that might impose significant costs on regulated entities? Are there ways to reduce these costs? What are the best ways to consider costs in relation to benefits?

29. How does the dearth of measurable standards or benchmarks impact the uptake of audits and assessments?

AI Accountability Policies

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?
    b. Should AI accountability regulation, if any, focus on inputs to audits or assessments (*e.g.*, documentation, data management, testing and validation), on increasing access to AI systems for auditors and researchers, on mandating accountability measures, and/or on some other aspect of the accountability ecosystem?
    c. If a federal law focused on AI systems is desirable, what provisions would be particularly important to include? Which agency or agencies should be

responsible for enforcing such a law, and what resources would they need to be successful?

d. What accountability practices should government (at any level) itself mandate for the AI systems the government uses?

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

32. What kinds of incentives should government explore to promote the use of AI accountability measures?

33. How can government work with the private sector to incentivize the best documentation practices?

34. Is it important that there be uniformity of AI accountability requirements and/or practices across the United States? Across global jurisdictions? If so, is it important only within a sector or across sectors? What is the best way to achieve it? Alternatively, is harmonization or interoperability sufficient and what is the best way to achieve that?

Dated: April 7, 2023

Stephanie Weiner

Acting Chief Counsel, National Telecommunications and Information Administration