2023

Comments of the Cordell Institute on AI Accountability

Neil M. Richards
Washington University in St. Louis School of Law, nrichards@wustl.edu

Woodrow Hartzog
Boston University School of Law

Jordan Francis

Repository Citation
Richards, Neil M.; Hartzog, Woodrow; and Francis, Jordan, "Comments of the Cordell Institute on AI Accountability" (2023). *Scholarship@WashULaw*. 499.
https://openscholarship.wustl.edu/law_scholarship/499

The Cordell Institute for Policy in Medicine & Law
Washington University in St. Louis
Anheuser-Busch Hall, Room 541B
St. Louis, Missouri 63130

June 12, 2023

National Telecommunications and Information Administration
1401 Constitution Avenue NW
Washington, DC 20230

Submitted via https://www.regulations.gov

Re: AI Accountability Policy Request for Comment, NTIA–2023–0005, (Docket No. 230407-0093)

Dear Administrator Davidson and NTIA Staff:

Thank you for the opportunity to comment on NTIA's recent inquiry into "what policies will help business, government, and the public be able to trust that Artificial Intelligence (AI) systems work as claimed—and without causing harm."[1] We applaud NTIA for taking this initiative and advancing the conversation about what rules and regulations will foster the creation of trust-preserving and trust-promoting AI systems. By centering this inquiry on the concepts of trust and risk of harm, NTIA is asking the right questions. We have long argued that trust and relational vulnerability are the critical lenses through which to view issues of privacy, data protection, and civil rights in the digital age.[2] This holds equally true for the design and implementation of AI systems. The mass adoption of AI systems exposes people as individuals and collectively to risks of harm because these systems are fueled by our personal data and, increasingly, are making consequential decisions about us in realms such as housing, employment, access to government benefits, healthcare, incarceration, and more. Promoting trustworthy AI therefore requires understanding the power disparities that exist between those who design and implement AI systems and the vulnerable, trusting humans subjected to these systems.

[1] Press Release, Nat'l Telecomm. & Info. Admin., AI Accountability Announcement and Panel Discussion (Apr. 11, 2023), https://ntia.gov/issues/artificial-intelligence/announcement-event.

[2] *See, e.g.*, Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431 (2016); Neil Richards & Woodrow Hartzog, *A Relational Turn for Data Protection?*, 6 EUR. DATA PROT. L. REV. 492 (2020).

We write to you today to argue that establishing trustworthy and accountable AI systems requires avoiding what we call "AI half-measures." The request for comments centers on the concept of "accountability mechanisms," which are often treated as being synonymous with audits, assessments, certifications, and similar procedural compliance requirements.[3] To that end, NTIA has asked, "[c]an AI accountability practices have meaningful impact in the absence of legal standards and enforceable risk thresholds?" The answer to that question is "no," because legal standards and enforceable risk thresholds[4] *are* the meaningful AI accountability practices because they place meaningful incentives on organizations to act in accountable ways. Similarly, NTIA's objectives include identifying "how supposed accountability measures might mask or minimize AI risks, . . . and ways governmental and non-governmental actions might support and enforce AI accountability practices."[5] Audits, assessments, and certifications might engender some trust in individuals who use or are subject to AI systems, and those might be helpful tools on the road to fostering trust and accountability. But there is no guarantee that trust created in this way will be reliable. Mere procedural tools of this sort will fail to create meaningful trust and accountability without a backdrop of strong, enforceable consumer and civil rights protections. (Q1(e).)

By contrast, legal standards and enforceable risk thresholds are substantive rules that offer a much better chance of successfully implementing AI systems within human-protective guardrails. Recognizing the importance of such substantive rules, we offer two arguments. First, a myopic focus on concepts of transparency, bias mitigation, and ethics (for which procedural compliance efforts such as audits, assessments, and certifications are proxies) is insufficient when it comes to the design and implementation of *accountable* AI systems. This is especially true where dangerous, disruptive systems are being released on the world by for-profit companies with little regard to the externalities or larger societal effects produced by these systems. We call rules built around transparency and bias mitigation "AI half-measures," because they provide the appearance of governance but fail (when deployed in isolation) to promote human values or hold liable those who create and deploy AI systems that cause harm. Second, any rules and regulations concerning AI systems must focus on substantive interventions rather than mere procedure. Flexible consumer protection standards, such as prohibitions on unfair, deceptive, and abusive acts or practices, are the kind of technology neutral measures which will protect individuals from harmful or unreasonably risky deployments of AI systems and encourage responsible innovation. Woven together as a vast regulatory fabric, these principles can invigorate and strengthen procedural tools such as audits and certifications, to the benefit of consumers both individually and as a group.

[3] *See* AI Accountability Policy Request for Comment, 88 Fed. Reg. 22,433, 22,435 (Apr. 13, 2023) ("Governments around the world, and within the United States, are beginning to require accountability mechanisms including audits and assessments of AI systems").

[4] *See generally* Margot Kaminski, *Regulating the Risks of AI*, 103 B. U. L. REV. (forthcoming 2023), https://papers.ssrn.com/abstract_id=4195066 (explaining the consequences of constructing AI harms as "risks" and comparing proposed and recently enacted AI risk regulation regimes).

[5] AI Accountability Policy Request for Comment, 88 Fed. Reg. at 22,435.

I. AI Half-Measures Are a Recipe for Failure

AI systems are best understood as what technology scholars call "socio-technical systems": They exist as complex assemblages of human and non-human actors at the intersection of many different forms of social, economic, and political life, shaped meaningfully by hardware and software as well as culture, social systems, economics, and legal rules.[6] These complex systems hold both great promise and great peril, and the purpose of AI accountability mechanisms is to maximize the individual and social benefits of these systems while minimizing their individual and social harms. This includes ensuring that our fundamental rights and freedoms are protected as we increasingly integrate AI into every aspect of our lives, whether we choose to as individuals or not. The widespread adoption of AI systems implicates many of the rights and principles we hold most dear: due process, freedom of expression, anti-discrimination, privacy, identity formation, the formation of meaningful and intimate social relationships, and opportunities for meaningful safe, healthy, and fulfilling work. It is imperative that we have accountability mechanisms in place to safeguard these rights and values and to forestall and remedy individual and social harms inflicted by untrustworthy AI systems (Q1.)

Governance of AI systems to foster trust and accountability requires avoiding the seductive appeal of "AI half-measures"—those regulatory tools and mechanisms like transparency requirements, checks for bias, and other procedural requirements that are necessary but not sufficient for true accountability. When implemented as standalone protections rather than components of broader governance strategies, AI half-measures provide only a veneer of accountability while failing to prevent or remedy the more serious harms that flow from deployment of untrustworthy AI systems. In so doing, AI half-measures reveal themselves as pernicious—offering the illusion of protection while enabling the festering of harms and other social costs. This makes AI half-measures appealing from an industry perspective but dangerous for society. In these comments we identify three such AI half-measures which are necessary but not sufficient for true accountability: transparency, bias mitigation, and ethics. These half-measures are the underlying goals that procedural tools like audits, assessments, and certifications seek to advance. Transparency, bias mitigation, and ethics are important and worthwhile goals. But they are not enough. We explain below how a myopic focus on these concepts will fail to sufficiently safeguard the interests of individuals who are using or subject to these systems.

Transparency. The broader conversation around AI governance and accountability is often dominated by discussion of the "black box" problem of AI and calls for increased transparency into these systems. The importance of transparency in AI accountability, however, depends wholly

[6] Mike Ananny & Kate Crawford, *Seeing Without Knowing: Limitations of the Transparency Ideal and Its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC'Y 973, 974, 983 (2018), available at https://doi.org/10.1177/1461444816676645.

upon what we mean by transparency and what it gets us—and particularly to what extent transparency furthers our ability to prevent and remedy harmful deployments of AI systems. To protect ourselves from the individual and social harms stemming from untrustworthy AI systems, we must do more than merely see into these systems; we must be capable of understanding them as assemblages and changing them when they do not align with our values. At best, transparency can only be a first step, and not as an end in itself.

Transparency means different things to different actors in the complex assemblages that constitute AI systems. Information about people, places, and things—which is collected through sensors embedded in Internet of Things devices, cell phones, click patterns, browsing history, social media activity, and direct input by individuals—are crucial inputs for the development and use of AI systems. Precipitated by rise of "big data analytics" in recent decades, the mass adoption of AI systems is premised on creating a more transparent world. Despite this, AI systems are anything but transparent themselves. Kate Crawford has provided detailed accounts about the "vast matrix of capacities" which are invoked whenever an individual interacts with an AI system, noting that "[t]he scale of this system is almost beyond human imagining."[7] We have previously labeled this internal conflict between the transparency AI promises and the opacity it is built upon as the Transparency Paradox of Big Data: "Big data promises to use this data to make the world more transparent, but its collection is invisible, and its tools and techniques are opaque, shrouded by layers of physical, legal, and technical privacy by design."[8] There are legitimate competitive and data security concerns underlying some of this secrecy, but much of this secrecy is unnecessary and harmful. As AI systems are increasingly relied upon to make predictions and consequential decisions concerning people, those people affected have a right to know the basis upon which those decisions are being made.[9] With the mass implementation of AI systems, we need "technological due process" that provides meaningful notice and transparency.[10] Systems which rely upon secretive surveillance or where decisions are made about people by a "Kafkaesque system of opaque and unreviewable decision-makers" cannot by definition be trustworthy.[11] It is critical, therefore, that we have some level of insight into the design and implementation of these systems, be that through certifications, audits, or assessments.

The need for transparency, however, should not predominate over the conversation about what policies and rules are necessary to craft trustworthy AI. Although transparency is a necessary condition for trustworthy AI, a myopic focus on transparency can come at the cost of deeper

[7] Kate Crawford & Vladan Joler, Anatomy of an AI System (2018), available at https://anatomyof.ai; *accord* Kate Crawford, Atlas of AI (2021).
[8] Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 42–43 (2013).
[9] *Id.* at 43.
[10] *Id.* (citing Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008)).
[11] *Id.*

engagement with the substantive applications of AI and the threats posed to people.[12] Kate Crawford and Mike Ananny have identified ten limitations of the transparency ideal:

1. It does not always follow that the ability to see inside a system results in the power to govern it.[13] If there are not systems in place to process, digest, and use the information revealed to create change, or if the decision-makers are not vulnerable to public exposure, then transparency does not result in meaningful change.[14]
2. Full transparency can be harmful, especially to vulnerable individuals.[15]
3. Transparency can intentionally occlude by making so much information available that it conceals more damaging information.[16]
4. Transparency can create false choices between complete secrecy and total openness if there is not a "nuanced understanding[] of the kind of accountability that visibility is designed to create."[17]
5. Transparency burdens individuals by forcing them to "seek out information about a system, to interpret that information, and determine its significance."[18] Similarly, the transparency ideal also presumes that different systems can easily be compared, allowing individuals to assess and choose between alternative options.[19]
6. There is a dearth of empirical evidence that transparency engenders trust, either trust in organizations and systems or by the organizations and systems making disclosures.[20]
7. Transparency is reliant on professionals who may have their own aims, such as "protecting the exclusivity of their expertise" or are subject to capture.[21]
8. Transparency efforts can prevent deeper understanding of complex systems by focusing on merely seeing into those systems rather than interacting with them more deeply.[22]
9. Technical limitations—resulting from the scale and speed of a system's design—can make a system inscrutable, even to its creators.[23] This problem is especially challenging in the context of machine learning AI systems such as deep learning.[24]

[12] Ananny & Crawford, *supra* note 6, at 974.
[13] *See id.* at 978.
[14] *Id.*
[15] *Id.* at 978–79; *see also* Richards & King, *supra* note 8, at 43 (noting that there are "legitimate arguments for some level of big data secrecy," such as protecting intellectual property rights).
[16] Ananny & Crawford, *supra* note 6, at 979.
[17] *Id.*
[18] *Id.* at 979–80.
[19] *Id.*
[20] *Id.* at 980.
[21] *Id.*
[22] *Id.* at 980–81.
[23] *Id.* at 981.
[24] *Id.*

10. Temporal limitations (i.e., whether transparency should mean "future relevance, anticipated revelation, ongoing disclosure, or *post hoc* visibility") alter the efficacy of transparency obligations because visibility at different moments in an AI system's lifetime may "require or produce different kinds of system accountability."[25]

In short, looking does not necessarily lead to knowing.[26] Nor does it produce accountability on its own. Given the complexities of algorithmic systems, "if accountability requires seeing a system well enough to understand it . . . using transparency for accountability begs the question of what, exactly, is being held to account."[27]

Other criticisms of the transparency ideal have emerged in recent years. In her recent article, *Algorithmic Grey Holes*, Professor Alicia Solow-Niederman also argues that AI accountability requires more than transparency, noting that "[a]lgorithmic grey holes can occur when layers of procedure offer a bare appearance of legality, without accounting for whether legal remedies are in fact available to affected populations."[28] Professor Solow-Niederman's argument focused on state deployment of algorithmic decision-making, but the underlying concern about a lack of redress applies to private-sector deployment of AI systems as well.

Bias Mitigation. The harms resulting from biased AI systems have been well-documented.[29] A common response to the bias problem is a call for procedural protections to mitigate bias. For example, the Federal Trade Commission has emphasized that AI tools used for consumer lending be "empirically derived, demonstrably and statistically sound."[30] It is certainly a good thing for AI developers to reduce statistical bias and error resulting from data that is unrepresentative or suffers from other faults such as mislabeling or lack of integrity. But treating bias mitigation as sufficient to ensure accountability runs into two problems. First, what constitutes "fairness" in the context of AI systems is a hotly debated topic. It is easy to say that AI systems should not be biased; it is very difficult to find consensus on what that means and how to approach that goal. As Professor Solow-Niederman has identified, "the very choice of a mathematical definition of 'fairness' is a political one."[31] Leaving industry actors to define bias on an individual basis will result in myriad competing definitions. Many of these will be thin and self-serving, and all of them will further

[25] *Id.* at 982.
[26] *See id.*
[27] *Id.* at 982.
[28] Alicia Solow-Niederman, *Algorithmic Grey Holes*, 5 J. L. & INNOVATION 116, 118 (2023).
[29] *See, e.g.*, Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. (2018).
[30] Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NW. U. L. REV. 357, 420 (2022) (citing Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM'N (Apr. 8, 2020), https://www.ftc.gov/news-events/blogs/business-blog/2020/04/using-artificial-intelligence-algorithms (quoting 12 C.F.R. § 1002.2 (2018) (Regulation B))).
[31] *Id.* (citing Arvind Narayanan, *Tutorial: 21 Fairness Definitions and Their Politics*, YOUTUBE (Mar. 1, 2018), https://www.youtube.com/watch?v=jIXIuYdnyyk).

complicate the ability of the average person to understand or trust AI systems. Second, focusing on bias-correction sidesteps a more fundamental question of whether an AI system should be deployed at all, as even perfectly accurate systems can produce biased (in the colloquial sense) or discriminatory outcomes.[32] As Professor Solow-Niederman has argued, there is a difference between AI systems that have a discriminatory effect on individuals even when developed and trained with representative data (e.g., emotion-recognition technologies) and those systems which have the potential for discriminatory effects when trained on non-diverse datasets or otherwise fail to follow best practices in development.[33] While the latter might only require standards and safeguards that guide development, the former are better suited for bright-line rules and bans.[34]

Ethics. Incorporating ethics-focused self-scrutiny into organizational culture through things like research on bias mitigation, dedicated ethics boards, and documented ethics goals sounds like accountability in the abstract—ethics should certainly be front of mind for companies that design or deploy AI systems. But AI ethics become AI half-measures where these commitments to ethics are commitments in name only. If there is no way to hold companies accountable for failure to align with established ethics, either because their statements are too vague in substance or because ethics boards hold no decision-making power, then ethics are operating as an AI half-measure.[35] These problems are compounded when there is little transparency as to how these companies are defining ethics and who is making those decisions, and when ethics can be defined in self-serving ways (such as forbidding practices a particular firm does not engage in, while turning a blind eye to its own practices that are dubious but profitable).[36] There is also little to suggest that adopting such principles stops companies from developing AI systems that contribute to unethical causes.[37]

* * *

[32] NTIA recognized this possibility in the request for comment. *See* AI Accountability Policy Request for Comment, 88 Fed. Reg. at 22,438 ("In some contexts, *not* deploying AI systems at all wil be the means to achieve the stated goals.").

[33] Solow-Niederman, *supra* note 30, at 419.

[34] *Id.* 419–20.

[35] James Vincent, *The Problem with AI Ethics*, VERGE (Apr. 3, 2019, 10:47 AM) https://www.theverge.com/2019/4/3/18293410/ai-artificial-intelligence-ethics-boards-charters-problem-big-tech ("Academic Ben Wagner says tech's enthusiasm for ethics paraphernalia is just 'ethics washing,' a strategy to avoid government regulation. When researchers uncover new ways for technology to harm marginalized groups or infringe on civil liberties, tech companies can point to their boards and charters and say, 'Look, we're doing something.' It deflects criticism, and because the boards lack any power, it means the companies don't change.").

[36] *Id.*

[37] *Id.* (first citing George Joseph, *Inside the Video Surveillance Program IBM Built for Philippine Strongman Rodrigo Duterte*, INTERCEPT (Mar. 20, 2019, 9:35 AM), https://theintercept.com/2019/03/20/rodrigo-duterte-ibm-surveillance; then citing Shannon Liao, *Google Employees Aren't Convinced That Dragonfly Is Dead*, VERGE (Mar. 4, 2019, 1:30 PM), https://www.theverge.com/2019/3/4/18250285/google-dragonfly-censored-search-engine-code-dead-employees-doubt) (including alleged examples of unethical projects by companies that had publicly committed to AI ethics).

The shortcomings of AI half-measures identified above should not be read to mean that transparency, bias mitigation, and ethics by design are not worthwhile measures to pursue. Our analysis merely highlights the need for a comprehensive, multipronged approach that is sensitive to context and that features structural, substantive protections. For example, the transparency limitations identified by Ananny and Crawford merely point to the need for nuanced transparency rules that tie what transparency aims to reveal to how visibility will result in meaningful change. The relevant questions should be "what do we want to see, how will that seeing lead to understanding, and how will that understanding lead to meaningful change?"[38] All too often, analysis of this sort has missed the final, crucial, meaningful step of asking how seeing better will produce better outcomes in reality. Likewise, our broad critique of procedural protections does not mean that procedural protections are not worth implementing; rather, they must be backed by substantive rules that promote human flourishing. Lawmakers cannot rely on procedural protections to duck the difficult question of determining what human values and goals constitute fair, trustworthy, and accountable AI.[39] As we have argued elsewhere, this approach has been tried in the privacy context for the last twenty-five years, and it has been a spectacular failure.[40] To avoid repeating such past mistakes, AI accountability measures should consider AI systems from a relational perspective. As Crawford and Ananny argue, AI systems are assemblages of human and non-human actors.[41] Understanding AI systems and ultimately holding them accountable requires understanding the relationships which underpin these systems and where the different components of these systems (algorithms, code, platforms, people, etc.) intersect.[42]

II. Substantive Protections and Structural Change

Industry leaders are quick to embrace AI half-measures, touting the benefits of "responsible AI" and the self-regulation measures which they believe will ensure that AI systems serve individuals and society.[43] This is a good thing in that developing trustworthy and accountable AI systems will require industry buy-in and cooperation. But lawmakers should not look at the efforts of a "small

[38] Ananny & Crawford, *supra* note 6, at 985 ("For any sociotechnical system, ask, 'what is being looked at, what good comes from seeing it, and what are we not able to see?'").

[39] *See* Solow-Niederman, *supra* note 30, at 421 ("Attention to the subject–processor leg of the triangle underscores the human beings affected by the act of information processing and foregrounds why process alone cannot answer the substantive question of what is 'unfair' here.").

[40] *See e.g.*, Neil M. Richards, Woodrow Hartzog & Jordan Francis, Comments of the Cordell Institute on the Prevalence of Commercial Surveillance and Data Security Practices that Harm Consumers, (Nov. 21, 2022), available at https://www.regulations.gov/comment/FTC-2022-0053-1071 (Comment ID: FTC-2022-0053-1071).

[41] Ananny & Crawford, *supra* note 6, at 983.

[42] *Id.*

[43] *See, e.g.*, Kolawole Samuel Adebayo, *Executives from Leading Companies Share How to Achieve Responsible AI*, FAST CO. (May 8, 2023), https://www.fastcompany.com/90891982/executives-from-leading-ai-companies-share-how-to-achieve-responsible-ai (sharing insights from industry leaders about how they are achieving "responsible AI" through internal governance policies).

cadre" of good actors and conclude that no further action is needed.[44] The fact that a small number of companies are engaging in "responsible AI" development is not sufficient. Rather than letting individual actors determine what "bias" is and how to mitigate it,[45] or whether techniques like anomaly-detection "solve" the "black box" problem of AI systems,[46] creating trustworthy and accountable AI systems will require a multipronged approach of procedural protections, flexible legal standards, and deep structural change. As discussed above, procedural protections like audits, assessments, and certifications that result in meaningful transparency, bias mitigation, and incorporation of ethics in design and deployment are necessary but not sufficient measures. This part of our letter instead focuses on the importance of flexible legal standards and structural change. (Q1(d).)

Flexible Standards. The conventional wisdom in tech policy that law cannot keep pace with technology is a gross misrepresentation. While rules may struggle to keep pace with technological innovation, standards have proven flexible and adaptable across time and technologies. The Federal Trade Commission's approach to promoting responsible innovation while still protecting Americans from unfair and deceptive practices is illustrative of how flexible standards can be responsive to new technologies, and this approach is a critical element of the kind of accountability needed to build trustworthy AI systems.

In her speech at the 2022 IAPP Global Privacy Summit, FTC Chair Lina Khan wisely called for lawmakers to pursue substantive, rather than procedural, privacy protections for consumers.[47] As generative AI has captured the world's imagination, Chair Khan has reiterated her commitment to pursuing substantive rules, proclaiming that the FTC will vigorously enforce the law against business models that exploit individuals.[48] Basic principles of consumer protection law have proven extremely durable and responsive to changes in society, the economy, and technology.

[44] *See id.* (describing a recent study showing that "a 'small cadre' of companies that are proactively pursuing responsible AI policies are also generation 50% more revenue growth than their peers").

[45] *See, e.g., id.*; *see also* Solow-Niederman, *supra* note 30, at 420 (citing Arvind Narayanan, *Tutorial: 21 Fairness Definitions and Their Politics*, YOUTUBE (Mar. 1, 2018), https://www.youtube.com/watch?v=jIXIuYdnyyk) (discussing how bias is a contested term in the context of AI systems).

[46] *See, e.g.*,, Adebayo, *supra* note 43.

[47] Lina Khan, *Fed. Trade Comm'n, Remarks of Chair Lina M. Khan as Prepared for Delivery IAPP Global Privacy Summit 2022 Washington D.C.*, FED. TRADE COMM'N (Apr. 11, 2022), https://www.ftc.gov/system/files/ftc_gov/pdf/Remarks%20of%20Chair%20Lina%20M.%20Khan%20at%20IAPP%20Global%20Privacy%20Summit%202022%20-%20Final%20Version.pdf (citing Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C. L. REV. 1687, 1693 (2020)) ("Going forward, I believe we should approach data privacy and security protections by considering substantive limits rather than just procedural protections, which tend to create process requirements while sidestepping more fundamental questions about whether certain types of data collection and processing should be permitted in the first place.")

[48] Lina M. Khan, *Lina Khan: We Must Regulate A.I. Here's How.*, N.Y. TIMES (May 3, 2023), https://www.nytimes.com/2023/05/03/opinion/ai-lina-khan-ftc-technology.html?smid=nytcore-ios-share&referringSource=articleShare.

Since unfair methods of competition were outlawed in 1914, consumer protection law has seen powerful new principles arise, including the FTC's power to prevent and remedy unfair and deceptive acts and practices as well as the more modern prohibition on abusive acts or practices. Despite being nearly a century old, these standards-based tools have retained remarkable flexibility to deal with consumer protection problems like false advertising, privacy policies, and weak data security that may have been unimaginable when these tools were created. In our own work we have argued that a duty of loyalty adapted from fiduciary law provides another such flexible, value-laden tool which policymakers can use to regulate the use of personal information as well as the design of digital tools.[49] These kind of flexible legal standards can enable accountability measures to be responsive to specific contexts and accommodate unknowns about downstream implementation. (Q15(d).)

In developing its report on what regulatory measures will assure external stakeholders that AI systems are legal, effective, ethical, safe, and otherwise trustworthy, NTIA should take note of the FTC's commitment to substantive legal protections for individuals using or subject to AI systems. The FTC's body of enforcement actions demonstrate that flexible legal standards can protect consumers without unduly stifling innovation or emergent technologies. Through these standards, the government should incentivize companies developing and implementing AI systems to adopt AI accountability measures by holding those companies liable for the harms they induce. Products liability law is not optional for companies; AI accountability measures should not be either. Engineers engage in stress testing. Drug and medical device manufacturers must navigate the FDA approval process. AI systems should be subject to similar testing before their release on an unsuspecting public,[50] and AI companies should be held responsible (and liable) for foreseeable harms to individuals and society. AI accountability mechanisms must consider the design and implementation of AI systems, including (1) the relative costs and benefits that flow to those creating AI systems and those affected by AI systems; and (2) the risks of harm that result from the use of these systems as well as the collection, processing, and transfer of personal data necessary for these systems to function. Through the application of flexible legal standards such as unfairness, deception, abusiveness, negligence, and loyalty, accountability mechanisms can remain sensitive to context and unknowns in downstream deployment and as time passes and technological and social contexts evolve. (Q1(a), Q1(d), Q30(c), Q32.)

Structural Change. Fostering trust and accountability in the AI ecosystem will require deep structural change. At the outset, policymakers should be aware of the competing versions of risk regulation (such as quantitative risk regulation, risk regulation as democratic oversight, allocating regulatory resources by risk, and enterprise risk management) and apply risk management tools

[49] *See generally* Woodrow Hartzog & Neil Richards, *The Surprising Virtues of Data Loyalty*, 71 EMORY L.J. 985 (2022).

[50] This is not necessarily an endorsement of an AI license system. Rather, it is a general argument in favor of meaningful, iterative testing through the lifecycle of AI systems, including pre- and post-market phases.

such as audits, assessments, and certifications with the goal of changing internal processes rather than merely signaling trustworthiness to consumers.[51] (Q2.) On a more substantive level, policymakers should consider vicarious liability and personal consequences for malfeasance by corporate executives. Again the FTC's enforcement actions concerning data privacy and data security are illustrative. The FTC has long held companies liable for providing the means and instrumentalities to unfair and deceptive conduct.[52] In a recent complaint filed against a geolocation data broker, the FTC alleged—among other things—that the sale of precise geolocation data "could enable third parties to track consumers' past movements to and from sensitive locations and, based on inferences arising from that information, inflict secondary harms including 'stigma, discrimination, physical violence, [and] emotional distress.'"[53] Holding companies liable for their conduct which enables secondary harms will be a critical tool in creating trustworthy and accountable AI systems. Another structural change that will be important for AI assurance is holding company executives personally liable for the harms which occur under their watch. In its recent enforcement action against online alcohol marketplace Drizly and its CEO James Cory Rellas, the FTC went beyond merely punishing Drizly for its data security failures and imposed continuing obligations on Rellas himself.[54] As the AI ecosystem develops, it will be important for consequences of harmful conduct to follow executives as they move between companies. It will also be important to develop meaningful individualized *ex post* process for individuals subject to AI systems, to compensate them for harms, protect dignity, and enhance the legitimacy of systems.

[51] *See* Kaminski, *supra* note 4, at 41–75 (detailing how risk regulation takes different forms—with different natures and goals—depending on how that form of risk regulation came to be and comparing four versions of risk regulation).

[52] *See, e.g.*, Remarks of Sheila F. Anthony, *13th Annual Advanced ALI-ABA Course of Study for In-House and Outside Counsel*, FED. TRADE COMM'N (Mar. 20, 1998), https://www.ftc.gov/news-events/news/speeches/advertising-unfair-competition-ftc-enforcement-0#N_4_ ("The 'means and instrumentalities' theory is a well established FTC legal principle.").

[53] Fed. Trade Comm'n v. Kochava Inc., No. 22-cv-00377, 2023 WL 3249808, at *6 (D. Idaho May 4, 2023) (quoting Complaint at ¶ 29) (describing the FTC's theory of harm as "plausible" but holding that the FTC failed to allege that consumers are suffering or are likely to suffer such secondary harms).

[54] *See* Order, Drizly, LLC & James Cory Rellas, FTC File No. 202-3185 https://www.ftc.gov/system/files/ftc_gov/pdf/202-3185-Drizly-Decision-and-Order.pdf.

* * *

The classic science fiction novel *Dune* employs a framing device by which each chapter is preceded by a quotation from a fictional book of wisdom, *Collected Sayings of Muad'Dib* by Princess Irulan. The *Dune* universe explores many themes concerning the rejection of thinking machines and conscious robots, and one of these epigrams is especially pertinent to our modern-day discussion of AI accountability: "The concept of progress acts as a protective mechanism to shield us from the terrors of the future."[55] The wisdom in this saying is not a rejection of technological advancement; rather, it is a call to action, reminding us to not let the abstract concept of progress be an excuse to avoid taking a hard look at the sometimes harmful realities of new technologies. In our work concerning trust and loyalty in the context of privacy law, we have written about the need for "a substantive embrace of a broad array of human values over privacy law's reflexive deference to individual choice, consent, and control. Lawmakers and industry love 'notice and choice' proceduralism because it allows them to avoid the difficult task of prioritizing human interests and making substantive interventions."[56] As AI systems are increasingly integrated into our lives, whether consumer-facing or operating in the background, there is an increased risk of harm being placed on the individuals whose personal data is being used to create these systems and the individuals to whom these systems may be applied to make consequential decisions. Now is the time for lawmakers to grapple with the difficult task of prioritizing human interests and making substantive interventions, rather than mechanically giving deference to industry deployments of AI systems under the guise of progress. This does not mean that lawmakers should reflexively ban AI technology *en masse*. To the contrary, lawmakers should encourage the design and implementation of AI systems in ways which embrace human values and promote human flourishing. They can achieve this through a multipronged approach of procedural protections, flexible legal standards, and deep structural change. But it does mean that invocation of magical terms like "progress" or "innovation" must no longer be used to abdicate regulatory responsibility over new technologies and business practices.

Although artificial intelligence has existed in one form or another for decades, the present moment is notable for the degree to which AI has captured the public imagination. ChatGPT and other public-facing AI systems dominate the headlines, and the latest wave of tech entrepreneurs are painting bold and imaginative visions of a future built upon AI. In developing its report on AI accountability policy development and the AI assurance ecosystem, NTIA should be equally bold and imaginative. As the commercial internet developed over the 1990s and beyond, the dominant regulatory approach was industry self-regulation, epitomized by a "notice and choice" regime which failed to safeguard individual or collective privacy.[57] As AI systems kickstart a new phase

[55] FRANK HERBERT, DUNE 410 (Penguin Books 2016) (1965).
[56] Hartzog & Richards, *supra* note 49, at 990.
[57] *See* MATTHEW CRAIN, PROFIT OVER PRIVACY: HOW SURVEILLANCE ADVERTISING CONQUERED THE INTERNET (2021).

of telecommunications and information systems revolution, we must avoid the mistakes of the past and proactively approach the difficult substantive issues raised by these technologies. Trust and accountability can only exist where the law provides meaningful protections for humans. And AI half-measures will certainly not be enough.

Respectfully submitted,

Neil Richards, Faculty Director[58]
Woodrow Hartzog, Fellow[59]
Jordan Francis, Legal Research Fellow[60]

[58] Koch Distinguished Professor in Law and Director, Cordell Institute, Washington University in St. Louis; Affiliated Fellow, Yale Information Society Project; Faculty Associate, Berkman Klein Center for Internet & Society at Harvard University; Affiliate Scholar, Stanford Law School Center for Internet & Society.

[59] Professor of Law, Boston University; Faculty Associate, Berkman Klein Center for Internet & Society at Harvard University; Affiliate Scholar, Stanford Law School Center for Internet & Society.

[60] Legal Research Fellow, 2022–23, Cordell Institute for Policy in Medicine & Law, Washington University in St. Louis.

Founded in 1853, Washington University in St. Louis (WashU) is an internationally known research university whose mission is to act in service of truth through the formation of leaders, the discovery of knowledge and the treatment of patients for the betterment of our region, our nation and our world. At WashU, we generate, disseminate, and apply knowledge. We foster freedom of inquiry and expression of ideas in our research, teaching and learning.

The Cordell Institute is a collaboration between the University's schools of law and medicine, founded to work on legal and other problems at the intersection of law and human information technologies. The Cordell Institute leadership and staff include includes law professors and attorneys who work at the forefront of privacy law, information law, and constitutional civil liberties. Its mission is to advance legal concepts toward a practical policy framework that guides the ethical use of human information to promote health, the protection of individuals, and education of the public. Our scholarship drives policies that reveal how human information can be used appropriately and effectively to build confidence and trust. The opinions offered in these comments are those of the undersigned members of the Cordell Institute in their scholarly capacities, and not necessarily those of Washington University as an institution or those of its constituent schools.