AGAINST AI HALF MEASURES

Woodrow Hartzog,[*] *Neil Richards,*[**] *Ryan Durrie*[***] *& Jordan Francis*[****]

ABSTRACT

So far, U.S. policy for artificial intelligence has largely consisted of industry-led approaches like encouraging transparency, mitigating bias, promoting principles of ethics, and empowering people. These approaches are vital, but they are only half measures. To bring AI within the rule of law, lawmakers must start drawing substantive lines.

In this essay, we identify four AI regulatory approaches as half measures. First, transparency does not produce accountability on its own. Next, while mitigating bias in AI systems is critical, even unbiased systems are a threat to the vulnerable. Third, while "AI ethics" are important, they are a poor substitute for laws. Finally, empowering people in their individual choices misses the larger questions about the distribution of power and collective wellbeing.

Instead of these half measures, we recommend that lawmakers reject the idea that AI systems are neutral and inevitable. When lawmakers go straight to putting up half-hearted guardrails, they fail to ask the existential question about whether some AI systems should exist at all. To avoid the mistakes of the past, lawmakers must make the hard calls. And AI half measures will certainly not be enough.

---

[*] Professor of Law, Boston University School of Law.

[**] 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.

[***] Associate Director, Cordell Institute for Policy in Medicine & Law, Washington University in St. Louis.

[****] Policy Counsel, Future of Privacy Forum; Fellow, Cordell Institute for Policy in Medicine & Law. The views expressed in this article are his own.

Table of Contents

Abstract 0
Introduction 0
I. Four Commonly Proposed AI Half Measures 5
A. Transparency 7
B. Incremental Technical Improvement 12
1. Bias Mitigation 12
2. PETs Magical Thinking 14
C. Self-Imposed Standards 15
D. Empowerment and Individual Control 19
II. Motivations for Half Measures: Resisting Neutrality, Inevitability, and Innovation Narratives 22
III. Substantive Interventions Beyond Half Measures 25
A. Flexible Duties of Loyalty and Care 28
1. Loyalty 29
2. Care 35
B. Specific Rules 37
1. Design Rules and Secondary Liability 37
2. Bright-line Prohibitions 39
3. Structural and *Ex Ante* Strategies 39
4. Private Right of Action and Hybrid Enforcement 42
Conclusion 43

INTRODUCTION

After years of the so-called "AI Winter," artificial intelligence systems are finally being deployed in large numbers. Consumer-facing AI tools like ChatGPT, CoPilot, Bing Assistant, and Midjourney are being used by millions of people for a wide range of personal tasks, while businesses are using a whole suite of AI tools for an even broader set of functions, such as logistics, HR, and other Taylorist efficiencies.[1] Notwithstanding all the overinflated AI hype, these technologies finally look capable of reshaping our society the way that the Internet transformed how we live, work, play, and participate in our democracy over the past three decades.[2] Yet all is not well with the AI Revolution. These systems have proven useful, but they also impose an enormous cost on our wellbeing, our environment, and our democracy. It is imperative that as AI deployment continues across the

[1] *See, e.g.*, IFEOMA AJUNWA, THE QUANTIFIED WORKER: LAW AND TECHNOLOGY IN THE MODERN WORKPLACE (2023).

[2] For more on the limits of AI, see, e.g., ARVIND NARAYANAN AND SAYASH KAPOOR, AI SNAKE OIL: WHAT ARTIFICIAL INTELLIGENCE CAN DO, WHAT IT CAN'T, AND HOW TO TELL THE DIFFERENCE (2024); MEREDITH BROUSSARD, ARTIFICIAL UNINTELLIGENCE: HOW COMPUTERS MISUNDERSTAND THE WORLD (2018).

range of economic, social, and political contexts that these technologies are adopted for reasons beyond making things cheaper.

Recognizing the challenges that AI poses to worker welfare, consumer protections, civil rights, and many other societal values, many companies, advocates, and reformers have proposed a variety of ways in which harms from AI technologies can be mitigated, whether through the form of technical fixes, the development of "AI ethics," or what are increasingly being referred to as "guardrails."[3] For example, President Biden's October 2023 Executive Order on AI calls for measures such as disclosure of reports and records relating to training of "dual-use foundation models,"[4] increased investment in privacy enhancing technologies (PETs),[5] and expanding global partnerships and collaborations on issues such as AI safety, testing, and transparency.[6] Although there is much to be lauded about the Biden Administration's whole-of-government approach to AI, these measures are starting points and bare-minimum "guardrails," rather than the kind of substantive, human-protective interventions necessary to ensure that AI systems are trustworthy and accountable.

The growing legislative and regulatory conversation surrounding AI also focuses on "accountability mechanisms," which are often treated as being synonymous with audits, assessments, certifications, and similar procedural compliance requirements. For example, in its June 2023 request for comments on AI accountability, NTIA asked, "Can AI accountability practices have meaningful impact in the absence of legal standards and enforceable risk thresholds?"[7] Similarly, one of NTIA's stated objectives in that RFC was identifying "how supposed accountability

---

[3] *See, e.g.*, URS GASSER AND VIKTOR MAYER-SCHÖNBERGER, GUARDRAILS: GUIDING HUMAN DECISIONS IN THE AGE OF AI (2024).

[4] Exec. Order No. 14,110, 88 Fed. Reg. 75,191, at 75,197–98 (Oct. 30, 2023).

[5] *Id.* 75,193.

[6] *Id.* 75,223–24.

[7] *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"). As we develop in this paper, we think the answer to that question is "no," because legal standards and enforceable risk thresholds *are* the meaningful AI accountability practices because they place meaningful incentives on organizations to act in accountable ways. *See generally* Margot Kaminski, *Regulating the Risks of AI*, 103 B. U. L. REV. 1347 (2023) (explaining the consequences of constructing AI harms as "risks" and comparing proposed and recently enacted AI risk regulation regimes).

measures might mask or minimize AI risks, . . . and ways governmental and non-governmental actions might support and enforce AI accountability practices."[8] These calls for reform generally encourage transparency, seek to mitigate bias, promote ethical principles, and aim to empower people using or facing these technologies. And that's the problem.

While these AI reform proposals are both important and well-meaning, existing AI reform proposals are unlikely to be enough to bring AI within the rule of law and to direct it to making society better rather than more efficient. Even more significantly, such proposals may fail to protect us while also giving the impression that our rules are sufficient and that lawmakers have done enough. In that sense these measures are best thought of as "AI half measures." The NTIA RFI reveals this when it asks whether AI Accountability principles can have meaningful effect absent legal standards and enforceable risk thresholds. It is difficult to consider something an accountability principle if it doesn't have meaningful effect.

Our argument in this essay is simple: *To bring AI within the rule of law, policymakers must go beyond half measures to ensure that AI systems and the actors that deploy them are worthy of our trust*. Trust and relational vulnerability are the critical lenses through which to view issues of privacy, data protection, and civil rights in the digital age.[9] The same is true for the design and implementation of AI systems. The mass adoption of AI systems exposes people—as individuals and collectives—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, access to essential goods and services, and more.[10]

---

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

[9] *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).

[10] *See, e.g.*, MEREDITH BROUSSARD, MORE THAN A GLITCH: CONFRONTING RACE, GENDER, AND ABILITY BIAS IN TECH (2023); KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE (2021); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); BRIAN CHRISTIAN, THE ALIGNMENT PROBLEM: MACHINE LEARNING AND HUMAN VALUES (2020); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2016); IFEOMA AJUNWA, THE

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.

We develop our claim in four parts. In Part I, we explain the concept of AI Half Measures, and argue how many of the most popular proposals for AI reform are half measures that may be worthy things to do, but which individually and collectively will not solve the varied challenges to our civilization that AI technologies pose. Four examples illustrate this point well.

First, we show how transparency proposals are insufficient because transparency does not automatically make things right when things go wrong. Second, proposals focused on incremental technological improvement—such as bias mitigation—are also insufficient because even if we ensure that AI works as intended and equally well for all communities, such an achievement will still create AI systems that can be used to dominate, damage, misinform, manipulate, and discriminate. Third, self-imposed standards such as ethical principles are important, but when ethical codes and advisory boards lack enforcement authority to change the design of technologies and the behavior of those who deploy them, they offer little more than lip service to the important limitations they are meant to impose, instead becoming "guardrail theater."[11] Fourth, while empowering individuals through control mechanisms or property rights in data may offer some limited protections in certain contexts, a focus on individuals would be mistaken when it comes to AI regulation because individual rights cannot solve structural, social problems. Thirty years of experience with modern privacy law makes one point clear – when it comes to complex technological systems, control over your informational destiny is not only an impossible illusion but it is one that can operate to make consumers do what the technology designers want them to do. Such commitments to control can have the opposite effect than what they are advertised as doing.

QUANTIFIED WORKER: LAW AND TECHNOLOGY IN THE MODERN WORKPLACE (2023); Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C.D. L. REV. 399 (2017); Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008); Daniel Solove, *Artificial Intelligence and Privacy*, 77 Fla. L. Rev. (forthcoming Jan 2025).

[11] Our thanks to Mary Madden, who offered this term during a session at the 2024 Privacy Law Scholars Conference.

Having stated the nature of the problem, the remainder of our article explains how we can move beyond AI Half Measures and use a variety of methods to bring AI within the rule of law. In Part II, we argue policymakers must reject the "Neutrality Fallacy" by accepting that AI is not neutral. This includes moving swiftly in holding developers of AI systems accountable for their design choices. We argue that policymakers should consider creating strong bright-line rules for the development and deployment of AI systems. For the most dangerous designs and deployments, lawmakers should impose outright prohibitions.

We also explain how policymakers must resist a related ideology to the Neutrality Fallacy, which we call the "Inevitability Narrative." Technologies like AI systems are not inevitable—they are intentionally designed and built by people, and people can prohibit them, they can regulate them, and they can shape their evolution into socially-beneficial tools as well. Lawmakers will make little progress until they accept that the toothpaste is never out of the tube when it comes to questioning and curtailing the design and deployment of AI systems for the betterment of society.

In Part III, we call for an increased focus on meaningful AI Half Measures, those substantive interventions that limit the abuses of power by complex sociotechnical systems like AI. Such approaches include imposing duties of loyalty, care, and confidentiality, design rules, bright line prohibitions, structural and ex ante approaches, and private causes of action.

Governance of AI systems to foster trust and accountability requires avoiding the seductive appeal of AI half measures. 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 presenting substantive interventions that move beyond half measures, this essay is not attempting to construct what would be the authors' ideal AI regulatory framework. Rather, the purpose is to encourage lawmakers who want to be proactive on shaping AI for good but don't know

where to start to start from a place of strength—rather than seductive but ultimately ineffective half measures—in their regulatory proposals. This is particularly important for state lawmakers who are often working on a litany of issues at once during shortened legislative sessions with little staffing or resources.

I. FOUR COMMONLY PROPOSED AI HALF MEASURES

It's clear that AI systems are going to change our world.[12] What's not clear yet is whether that change will be worth it.[13] These complex systems hold both great promise and great peril, and the ostensible purpose of AI accountability mechanisms is to maximize the individual and social benefits of these systems while minimizing their individual and social harms. The widespread adoption of AI systems implicates many foundational human rights and democratic values: 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. Most AI systems are by their nature hungry for data,[14] leaky at holding this data,[15] sneaky in how they gather and use data,[16] and exclusory in their results and consequences for

---

[12] We use the term AI systems here to reflect what Science and Technology Studies 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. *See* Ryan Calo, The Scale and the Reactor, https://papers.ssrn.com/abstract_id=4079851; Daniella DiPaola & Ryan Calo, Socio-Digital Vulnerability, https://papers.ssrn.com/abstract_id=4686874; 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.

[13] *See, e.g.*, Woodrow Hartzog, *Two AI Truths and a Lie*, 26 YALE J. L. & TECH (forthcoming 2024).

[14] Robert Hart, *Clearview AI Fined $9.4 Million In U.K. For Illegal Facial Recognition Database*, FORBES (May 23, 2022, 6:55am), https://www.forbes.com/sites/roberthart/2022/05/23/clearview-ai-fined-94-million-in-uk-for-illegal-facial-recognition-database/?sh=c9ef95019636; Alex Reisner, *These 183,000 Books Are Fueling The Biggest Fight In Publishing And Tech*, THE ATLANTIC (Sept. 25, 2023), https://www.theatlantic.com/technology/archive/2023/09/books3-database-generative-ai-training-copyright-infringement/675363/.

[15] Ben Derico, *ChatGPT bug leaked users' conversation histories*, BBC (Mar. 2023), https://www.bbc.com/news/technology-65047304; James Vincent, *Apple restricts employees from using ChatGPT over fear of data leaks*, THE VERGE (May 19, 2023, 3:29 AM), https://www.theverge.com/2023/5/19/23729619/apple-bans-chatgpt-openai-fears-data-leak.

[16] Woodrow Hartzog, *Unfair and Deceptive Robots*, 74 MAR. L. REV. 785 (2015).

consumers.[17] Data exploitation and manipulation, whether in the context of creating and deploying AI systems, comes from an imbalance of power and information in relationships.[18] Tech companies can control what their customers see and they collect information from across the web as though all accessible human information is theirs for the taking. Compared to individual consumers, tech companies have practically unlimited resources and strong financial incentives to influence people's behavior for their own benefit and profit.

Our need for meaningful AI regulation underscores the stakes of settling for half measures. To start, we conceive of AI half measures as an action or policy that is not forceful or decisive enough to respond to the individual and collective risks of AI systems. Half measures are often procedural in nature, narrow in scope, or lack meaningful enforcement mechanisms. They frequently take the form of post-deployment controls, audits, assessments, certifications, and similar procedural compliance requirements.[19] These tools are necessary to begin the task of data governance, but industry has routinely leveraged procedural checks such as these to dilute data and consumer protection law into a managerial box-checking exercise that largely serves to entrench harmful surveillance-based business models.[20] A checklist is no match for the staggering fortune available to those who exploit our data, labor, and precarity to develop and deploy AI systems. And it's no substitute from meaningful liability for when AI systems harm the public.[21]

---

[17] *See* Neil Richards & Woodrow Hartzog, *Against Engagement* (conference draft , Privacy Law Scholars Conference 2022) (on file with authors).

[18] *See, e.g.* Woodrow Hartzog & Neil Richards, *Legislating Data Loyalty*, 97 NOTRE DAME L. REV. REFLECTION 356 (2022).

[19] *See* 88 Fed. Reg. 22,433, 22,435, *AI Accountability Policy Request for Comment* (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").

[20] *See generally*, Ari Ezra Waldman, Industry Unbound: The Inside Story of Privacy, Data, and Corporate Power (2021).

[21] See, e.g., Catherine M. Sharkey, A Products Liability Framework for AI, 25 Columb. Sci. & Tech. L. Rev. (2024); Woodrow Hartzog, Unfair and Deceptive Robots, 74 Maryland Law Review 785 (2015).

The harms of AI are real, significant, and becoming both entrenched and normalized by the day.[22] But lawmakers have also sought to preserve the benefits of AI in while minimizing its cost to people and society. The result is that they have turned to predictable and popular strategies for regulating problems related to information technology—the "ol' faithful" approaches of transparency and control combined with a new focus on mitigating wrongful bias in systems and promoting self-regulatory approaches to keep companies ethical. Let's take each one in turn.

A. 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. But looking into a system does not necessarily lead to knowing about it.[23] Nor does it produce accountability on its own. Transparency must work in tandem with due process and additional, substantive legal mechanisms that ensure people are not stripped of their rights.

The importance of transparency in AI accountability depends wholly 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 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,

---

[22] *See generally*, Grant Ferguson et al., *Generating Harms: Generative AI's Impact & Path's Forward*, Electronic Privacy Information Center (May 2023), https://epic.org/wp-content/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf; Woodrow Hartzog, Evan Selinger & Johanna Gunawan, *Privacy Nicks: How the Law Normalizes Surveillance*, 101 Wash. U. Law Rev. 717 (2024); Hartzog, *supra* note 13.

[23] Mike Ananny & Kate Crawford, *Seeing Without knowing: Limitations of the Transparency Ideal and its Application to Algorithmic Accountability*, 20 NEW MEDIA & SOC'Y 973, 977-982 (2018).

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."[24]

Neil Richards and Jonathan King 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."[25] 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.[26] With the mass implementation of AI systems, we need "technological due process" that provides meaningful notice and transparency.[27] 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.[28] 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

---

[24] Kate Crawford & Vladan Joler, Anatomy of an AI System (2018), available at https://anatomyof.ai; *accord* Kate Crawford, Atlas of AI (2021).

[25] Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 42–43 (2013).

[26] *Id.* at 43.

[27] *Id.* (citing Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008)).

[28] *Id.*

deeper engagement with the substantive applications of AI and the threats posed to people.[29] 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.[30] 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.[31]
2. Full transparency can be harmful, especially to vulnerable individuals.[32]
3. Transparency can intentionally occlude by making so much information available that it conceals more damaging information.[33]
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."[34]
5. Transparency burdens individuals by forcing them to "seek out information about a system, to interpret that information, and determine its significance."[35] Similarly, the transparency ideal also presumes that different systems can easily be compared, allowing individuals to assess and choose between alternative options.[36]
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.[37]

---

[29] Ananny & Crawford, *supra* note 12, at 974.
[30] *See id.* at 978.
[31] *Id.*
[32] *Id.* at 978–79; *see also* Richards & King, *supra* note 25, at 43 (noting that there are "legitimate arguments for some level of big data secrecy," such as protecting intellectual property rights).
[33] Ananny & Crawford, *supra* note 12, at 979.
[34] *Id.*
[35] *Id.* at 979–80.
[36] *Id.*
[37] *Id.* at 980.

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.[38]
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.[39]
9. Technical limitations—resulting from the scale and speed of a system's design—can make a system inscrutable, even to its creators.[40] This problem is especially challenging in the context of machine learning AI systems such as deep learning.[41]
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."[42]

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."[43]

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."[44] Professor Solow-Niederman's argument focused on state deployment of algorithmic decision-making, but the underlying concern about a lack of

---

[38] *Id.*
[39] *Id.* at 980–81.
[40] *Id.* at 981.
[41] *Id.*
[42] *Id.* at 982.
[43] *Id.* at 982.
[44] Alicia Solow-Niederman, *Algorithmic Grey Holes*, 5 J. L. & INNOVATION 116, 118 (2023).

redress applies to private-sector deployment of AI systems as well. Considering transparency's value in light of its limitations reveals transparency to be an AI Half Measure, albeit a foundational and stackable one.

A survey of recently proposed legislation at the federal and state levels illustrates the degree to which policymakers are focusing on making AI systems transparent. These proposals vary wildly in scope and intention. Some bills would require developers or deployers of AI systems to make certain disclosures to the State, likely so as to inform future legislation or enforcement of existing laws.[45] Another category of bills require deployers of AI tools to provide information directly to individuals either using the AI system or subject to it.[46] There are innumerable generative-AI disclosure, watermarking, and provenance bills, especially within the context of political advertising.[47] Many of these bills represent sound policy decisions and should be seriously pursued by lawmakers. But these proposals would offer little protection to individuals as standalone requirements independent of a broader AI regulatory scheme.

Notably, several major legislative and regulatory proposals include transparency or notice as a component of a broader set of regulatory interventions, including risk or impact assessments, AI governance program requirements, opt out rights, heightened ex-post rights of access to information about systems, and even flexible legal duties like a duty of care to avoid discrimination.[48] These broader proposals that pair

---

[45] *See, e.g.*, A.B. 3204, 2023–24 Cal. State Assembly, Reg. Sess. (Cal. 2024) (requiring "data digesters"—entities using personal data to train AI systems—to register with the state and provide documentation regarding the personal information used to train its AI systems).

[46] *See, e.g.*, A.B. 2013, 2023–24 Cal. State Assembly, Reg. Sess. (Cal. 2024) (requiring that developers of AI systems or services made available to the public for use publicly post documentation concerning the data used to train the system or service).

[47] For watermarking, authenticity, and content provenance bills, see, e.g., A.B. 3211, 2023–24 Cal. State Assembly, Reg. Sess. (Cal. 2024) (imposing varying transparency requirements on generative AI providers (creation of watermarks and provenance data), entities deploying conversational AI systems (notice that such systems are in use), and large online platforms (labeling synthetic content)); A.B. 3050, 2023–24 Cal. State Assembly, Reg. Sess. (Cal. 2024); S.B. 217, 135th Gen. Assembly, Reg. Sess. (Ohio 2024); Advisory for AI-Generated Content Act, S. 2765, 118th Cong. (2023).

[48] *See, e.g.*, S.B. 2, Conn. Gen. Assembly, Reg. Sess. (Conn. 2024) (a comprehensive, risk-based AI regulation that has developer and deployer obligations for

transparency with additional obligations and individual rights are laying the groundwork for what are AI full-measures.

B. Incremental Technical Improvement

Another category of commonly proposed AI Half Measures is incremental technical improvement. The logic is simple and seductive: If the AI system in question worked better, then issues of harm and inequity would be resolved. This is evident in the vocal criticisms that AI systems are biased. A less publicly prominent proposal is the increased investment in and adoption of privacy enhancing technologies (PETs). Both, though virtuous in their own way, are insufficient accountability principles on their own.

1. Bias Mitigation

AI systems are notoriously and perhaps inevitably biased. A host of scholars have spent decades identifying the ways in which AI systems are biased against marginalized and underrepresented communities, most notably along the familiar lines of race, class, gender, and ability.[49] To

high-risk AI systems, generative AI systems, and "general-purpose" AI models); AB 2930, 2023–24 Cal. State Assembly, Reg. Sess. (Cal. 2024); SB 24-205, 74th Gen. Assembly, Reg. Sess. (Colo. 2024).

[49] *See, e.g.*, IFEOMA AJUNWA, THE QUANTIFIED WORKER: LAW AND TECHNOLOGY IN THE MODERN WORKPLACE (2023); MEREDITH BROUSSARD, MORE THAN A GLITCH: CONFRONTING RACE, GENDER, AND ABILITY BIAS IN TECH (2023); SAFIA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018); RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE (2021); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018); CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); SIMONE BROWNE, DARK MATTERS: ON THE SURVEILLANCE OF BLACKNESS (2015); KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE (2021); Batya Friedman and Helen Nissenbaum, *Bias in Computer Systems*, 14 ACM TRANSACTIONS ON INFORMATION SYSTEMS 330 (July 1, 1996); Aylin Caliskan, Joanna J Bryson & Arvind Narayanan, *Semantics derived automatically from language corpora contain human-like biases*, SCIENCE (April 14, 2017) https://pubmed.ncbi.nlm.nih.gov/28408601/; Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove & Aaron Rieke, *Discrimination through Optimization: How Facebook's Ad Delivery can Lead to Biased Outcomes*, 3 PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION 1 (Nov. 7, 2019); Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. (2018); Ngozi Okedigbe, *Discredited Data*, 107 CORNELL L. REV. 2007 (2022).

mitigate bias, lawmakers, regulators, and those in industry call for more care in the development of AI systems.[50]

A variety of enacted and proposed bills at the state and federal levels discuss mitigating bias in systems but fail to address deeper bias in the way those systems get deployed on disadvantaged populations. For example, in October 2023, California enacted Assembly Bill 302, a law which requires the California Department of Technology to conduct an inventory of high-risk automated decision systems proposed for use or used by state agencies.[51] With respect to bias, the law requires the inventory to include a description of "measures in place, if any, to mitigate the risks, including . . . the risk of inaccurate, unfairly discriminatory, or biased decisions, of the automated decision system."[52] This provision, though an important consideration, is largely descriptive and falls short of even mandating that there be bias mitigation protections in place. A similar bill enacted in Indiana in 2024 requires state agencies to produce an inventory of AI technologies in use and include in that inventory a description of whether the technology's data or information output has been evaluated for and found to exhibit bias.[53] Similar bills have been proposed at the federal level,[54] and this focus on describing measures taken to mitigate bias also appear in bills regarding private sector provision and use of AI tools.[55]

Bias mitigation is undoubtedly worthy of attention, resources, and regulation. AI systems will remain dangerously and possibly fatally flawed

---

[50] *For example*, the Federal Trade Commission has emphasized that AI tools used for consumer lending be "empirically derived, demonstrably and statistically sound." 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))).

[51] California Assembly Bill 302, Reg. Sess. 2023–24, (Cal. 2023).

[52] *Id.*

[53] Indiana Senate Bill 150, Reg. Sess. 2024 (In. 2024).

[54] *E.g.*, Eliminating Bias in Algorithmic Systems Act of 2023, S. 3478, 118th Cong. (2023) (requiring federal agencies to submit reports to congressional committees detailing bias risks of covered algorithms, steps taken to mitigate harms stemming from bias, actions taken to engage with relevant stakeholders regarding bias, and "relevant recommendations for legislation or administrative action to mitigate bias").

[55] A bill introduced in New Jersey in 2024, S1588, would prohibit selling an automated employment decision tool unless that tool has been the subject of a "bias audit" in the prior year, which entails "an impartial evaluation" to assess the tool's predicted compliance with the state's laws regarding employment discrimination. S. 1588, Reg. Sess. 2024–25 (N.J. 2024).

so long as they reflect harmful societal discriminatory practices. However, while indispensable, by itself, de-biasing AI systems risks being a half measure in two ways.

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."[56] 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 complicate the ability of the average person to understand or trust AI systems.

Second, when lawmakers and industry focus on bias-correction, they seem to assume away important threshold questions about whether AI has virtuous goals and uses, whether in a particular context or more generally. It is tempting to think a less biased AI system is less dangerous. But accurate AI systems are, if anything, more dangerous.[57] Less-biased AI is more attractive to the powerful, who can abuse it. When industry uses bias mitigation to gloss over the other imbalanced dynamics of AI, they misrepresent the lessons from these important scholars teaching us that bias is the symptom of a larger problem about how power is amassed and wielded against marginalized communities.[58] To put this point simply, even if we ensure that AI works equally well for all communities, such an achievement will still create AI systems that can be used to dominate, damage, misinform, manipulate, and discriminate.

2. PETs Magical Thinking

Privacy enhancing technologies (PETs) are the topic du jour in the privacy community. President Biden's landmark October 2023 executive order on artificial intelligence calls for federal agencies to research,

---

[56] *Id.* (citing Arvind Narayanan, *Tutorial: 21 Fairness Definitions and Their Politics*, YOUTUBE (Mar. 1, 2018), https://www.youtube.com/watch?v=jIXIuYdnyyk).

[57] *See* Evan Selinger and Woodrow Hartzog, *What Happens When Employees Can Read Your Facial Expressions?*, NEW YORK TIMES (Oct. 17, 2019), https://www.nytimes.com/2019/10/17/opinion/facial-recognition-ban.html.

[58] *See generally,* Anita L. Allen, *Dismantling the "Black Opticon": Privacy, Race Equity, and Online Data-Protection Reform*, 132 YALE L.J. FORUM 907 (2022).

develop, and implement PETs.[59] Broadly speaking, PETs are technical tools which, if designed and implemented correctly, can enable socially beneficial data sharing and use while mitigating privacy risk to individuals that would normally stem from the disclosure of personal data.[60] Examples of PETs include "secure multiparty computation, homomorphic encryption, zero-knowledge proofs, federated learning, secure enclaves, differential privacy, and synthetic-data-generation tools."[61]

Where the primary risk to an individual is the inappropriate disclosure of their personal data, such as risks of confidentiality or integrity, PETs have considerable promise to protect individuals from privacy harms. But invocation of PETs alone will not protect individuals as if by magic. Like with bias mitigation, if the substantive purpose of an AI system is harmful, then no implementation of tools like on-device processing can properly protect the affected individuals.

* * *

Incremental technical improvements in the form of bias mitigation and utilization of PETs are laudable endeavors motivated by a desire for equity and to protect the humans who interact with AI systems. These are, therefore, foundational and stackable AI Half Measures, worthy of pursuit but required to exist alongside additional safeguards and substantive measures to ensure that technical improvements are not creating more-perfect tools of domination.

C. Self-Imposed Standards

Ethics should certainly be front of mind for companies that design or deploy AI systems. The good news is that ethics—whether embodied in principles, codes, and boards—is popular with industry. Scholars at the Berkman Klein Center for Internet & Society at Harvard University recognized that "[i]n the past several years, seemingly every organization with a connection to technology policy has authored or endorsed a set of

---

[59] Exec. Order No. 14,110, 88 Fed. Reg. 75,191 (Oct. 30, 2023).

[60] The AI Executive Order defines PETs as "any software or hardware solution, technical process, technique, or other technological means of mitigating privacy risks arising from data processing, including by enhancing predictability, manageability, disassociability, storage, security, and confidentiality." *Id.* at 75,195.

[61] *Id.*

principles for AI."[62] In their deep dive into the contents of thirty-six prominent AI principles documents, the scholars found eight key themes in the sea of documents, including privacy, accountability, safety and security, transparency and explainability, fairness and non-discrimination, human control of technology, professional responsibility, and promotion of human values.[63] Sounds good. And familiar. Most of the principles analyzed have roots in human rights law and the fair information practice principles.[64]

But ethical principles for AI become AI half measures where these commitments to ethics are commitments in name only, or when they happen after the fact once AI tools have been built.[65] Ethical principles operate as a half measure if there is no way to hold companies accountable for failure to align with their espoused commitments, either because their statements are too vague or self-serving in substance or because ethics boards hold no decision-making or accountability power.[66] When industry does no more than adopt ethical principles, we get easy-to-make promises by companies to avoid the practices that aren't in their business model, but silence regarding the dubious tools that can make them money. We also get

---

[62] Jessica Fjeld, Nele Achten, Hannah Hilligoss, Adam Christopher Nagy, Madhulika Srikumar, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482.

[63] Id.

[64] Id. ("64% of our documents contained a reference to human rights, and five documents [14%] took international human rights as a framework for their overall effort.").

[65] In fact many companies have already given up on their own commitments. Gerrit De Vynck

& Will Oremus, *As AI booms, tech firms are laying off their ethicists*, THE WASHINGTON POST (March 30, 3023), https://www.washingtonpost.com/technology/2023/03/30/tech-companies-cut-ai-ethics/.

[66] 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.").

companies continuing to develop dubious AI systems that violate their own ethics principles.[67]

Professor Ryan Calo has noted that we are flooded with ethical principles from industry, the government, and even civil society, but this flood has not been accompanied by substantive legal rules.[68] Ethical principles are a poor substitute for laws and can even delay eventual rules because espousing principles and pointing to ethics committees can give the illusion of progress. It's easy to publicly commit to ethics, but industry doesn't have the incentive to leave money on the table for the good of society.[69] Ethics are important, but unless they occur throughout the development of AI systems rather than just at the end, and unless they are accompanied by substantive, external, legal constraints and sanctions,

---

[67] *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); see also, Fjeld, et. al., supra note ^ ("Moreover, principles are a starting place for governance, not an end. On its own, a set of principles is unlikely to be more than gently persuasive. Its impact is likely to depend on how it is embedded in a larger governance ecosystem, including for instance relevant policies (e.g. AI national plans), laws, regulations, but also professional practices and everyday routines.").

[68] Ryan Calo, *Artificial Intelligence and the Carousel of Soft Law*, IEEE (Sept. 16, 2021), https://ieeexplore.ieee.org/document/9539878 ("[G]enerally speaking, there has been little change to the law and legal institutions in light of the supposedly transformative technology of our time. What we have instead are principles. We are awash in them. The industry has a running supply: Microsoft, Google, Facebook, IBM, and other companies have each released principles, despite joining with civil society to form an organization—the Partnership on AI—which has its own "tenets." The White House has AI principles. The Department of Defense has AI principles. So does the U.S. Chamber of Commerce. The UN, the WTO, and the OECD all have published AI principles. Even organizations with a track record of advocating for concrete legal reform have released or endorsed AI principles.").

[69] *Id.* ("The impulse of so many organizations across nearly every sector of society to promulgate principles in response to the ascendance of AI is understandable. Unlike law, which requires consensus and rigid process, an organization can develop and publish principles unilaterally…and while common principles can lay a foundation for societal change, they are no substitute for law and official policy … No invisible hand guides market participants to charity. The Internet is not Eden. Uber and Airbnb are not sharing with anyone. And AI is not a magical genie-in-training … The role of the law is to understand, channel, and address that change—with rules, not aspirations.").

ethics are a half measure.[70] For this reason, self-imposed standards such as codes of ethics are ineffective—rather than foundational and stackable—AI Half Measures.

The same can be said for the spate of bills that create a committee to research the effects of AI and propose regulatory interventions. It seems like common wisdom that committees are where political inertia goes to die. In 2017—a lifetime ago in the tech policy world—New York City passed a local law establishing a task force to examine how the City's use of "automated decision systems" could harm people. At the time, Julia Powles penned a prescient article in the *New Yorker* arguing that the fact-finding taskforce faced a "Sisyphean" task without meaningful authority to require disclosures from agencies and outside contractors and failed to properly make use of the City's leverage in dealings with AI developers, supporting AI developers at the expense of meaningful public accountability.[71] Since then, similar task forces have been suggested as a means of establishing AI accountability. In 2024 alone, dozens of such bills have been introduced which would establish AI ethics or safety committees, commissions, or task forces.[72] Fact-finding and evidence-based decision-making are critical for

---

[70] *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).

[71] Julia Powles, *New York City's Bold, Flawed Attempt to Make Algorithms Accountable*, NEW YORKER (Dec. 20, 2017), https://www.newyorker.com/tech/annals-of-technology/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable.

[72] *See, e.g.*, New York Senate Bill 8755 (2024) (establishing a state AI ethics commission, in addition to substantive requirements regarding discrimination, dissemination of false information, and more); Washington Senate Bill 5838 and House Bill 1934 (2024) (would establish an AI task force to make legislative recommendations for generative AI standards); Indiana Senate Bill 150 (2024) (creating a task force to study government use of AI and create a report with legislative recommendations); Oregon House Bill 4153 (2024) (creating a task force to study definitions of AI to use in future legislation); West Virginia House Bill 5690 (2024) (creating a task force to recommend AI policies for certain uses of AI); West Virginia House Bill 5490 (2024) (creating a task force to study use of generative AI in schools and public service); Illinois House Bill 3563 (2023) (establishing a task force to report on generative AI); Delaware House Bill 333 (2024) (creating an AI commission to make recommendations as to how to ensure that government use of AI is safe and not rights-violating); Maryland Senate Bill 1087 (2024) (creating an AI commission to study AI use and report recommendations); Rhode Island House Bill 7158 (creating a

passing effective legislation, but such efforts must be more than acts of vanity or opportunities for credit-claiming with respect to "acting on AI."

These kinds of task forces can be helpful. At least one such committee has directly lead to substantive AI legislation being introduced: Connecticut Senate Bill 2, which passed the Connecticut Senate in April 2024 and would have regulated the use of "high-risk" AI systems, was in-part the product of the Connecticut AI Working Group.[73] But that success story is proving to be the exception rather than the rule.

D. Empowerment and Individual Control

One of the most popular approaches to legislating AI (as with privacy and big data before it) has been to seek to give individuals more control or ownership over their own data and creations. We can see this in new bills that require consent for data practices, give people rights over their data, and seek to give people intellectual property rights in their artistic creations and their names and likenesses. But when legislation becomes too focused on the individual at the expense of society, it can become a half measure for two reasons. First, strategies to give people more control can backfire, because it's easy to overwhelm people with choices and delude them about what's really going on. Second, focusing on the collective wisdom of trillions of self-motivated, nudgeable, and possibly misinformed decisions isn't always what's best for society.

---

commission to study use of AI and report legislative recommendations); Tennessee Senate Bill 1651 (directing an existing commission to study use of AI and report legislative recommendations); Virginia Senate Bill 487 (creating a commission to create recommendations for regulating AI, including ethical principles). *See generally* Lawrence Norden & Benjamin Lerude, *States Take the Lead on Regulating Artificial Intelligence*, BRENNAN CENTER (Nov. 6, 2023), https://www.brennancenter.org/our-work/research-reports/states-take-lead-regulating-artificial-intelligence (noting that, at the time of writing, at least 12 states had enacted laws establishing bodies such as task forces, advisory boards, commissions, committees, and councils for the purpose of studying AI and that such bodies can be a "first step toward regulatory action").

[73] *See* Zach Williams, *Connecticut Lawmaker Gains Prominence as States Grapple With AI*, BLOOMBERG NEWS (Feb. 14, 2024), https://news.bloomberglaw.com/artificial-intelligence/connecticut-lawmaker-gains-prominence-as-states-grapple-with-ai; Tatiana Rice, *Opinion: Will CT Seize Its Opportunity to lead in Responsible AI?*, CT MIRROR (May 6, 2024), https://ctmirror.org/2024/05/06/will-ct-seize-its-opportunity-to-lead-in-responsible-ai (praising SB 2 for its "rigorous development process" and multistakeholder engagement, which stemmed in-part from the state's 2023 establishment of an AI working group "to craft recommendations for private-sector AI use").

Thirty years of experience with modern privacy law makes one point clear – when it comes to complex technological systems, control over your informational destiny is not only an impossible illusion but it is one that can operate to make consumers do what the technology designers want them to do. Such commitments to control can have the opposite effect than what they are advertised as doing.[74]

Lawmakers who prioritize individual control, consent, and ownership in a vacuum risk missing how power and information are unequally distributed and deployed. Quite simply, the transparency and control contemplated by these frameworks is impossible in mediated environments. People can only click on the options provided to them and companies have incentive to design their products to nudge and manipulate people into doing what the designers want them to, whether through "dark patterns," hidden options, or "are you sure?" popups. Rules that prioritize individual control over data create incentives for companies to hide the risks of AI systems through manipulative design, vague abstractions, and complex or soothing words as they force us to accept those risks by designing systems where we never stop clicking the "I agree" button. This is, of course, assuming we had a choice about whether to use these systems in the first place. Lawmakers should not seek to give consumers dubious rights with which to try to protect themselves; lawmakers should seek to actually protect consumers regardless of what they choose.

Similarly, decisions about who owns the output of AI trained on or using other's data, works of art, or even likenesses are fundamental decisions for lawmakers. However, these decisions in and of themselves will not mitigate the effects of untrustworthy and unaccountable AI. Apportioning ownership of certain output of an AI that has already ingested and integrated a consumer's information or intellectual property may be able to redress some harms, but only to the extent that the infraction is identifiable and remediable by such an ownership transfer. Ownership is not the answer to the problem of accountability, trust, and equitable access to the powers of AI.

* * *

---

[74] *See* Neil Richards, Why Privacy Matters 90-100 (2022).

The shortcomings of AI half measures identified above should not be read to mean that transparency, bias mitigation, voluntary ethics codes, and individual control are not worthwhile measures to pursue. But we should not view them as end goals. 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?"[75] 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.[76] 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.[77]

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.[78] Understanding AI systems and ultimately holding them accountable requires understanding the relationships which underpin

---

[75] Ananny & Crawford, *supra* note 12, 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?'").

[76] *See* Solow-Niederman, *supra* note 50, 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.").

[77] 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).

[78] Ananny & Crawford, *supra* note 12, at 983.

these systems and where the different components of these systems (algorithms, code, platforms, people, etc.) intersect.[79]

II. MOTIVATIONS FOR HALF MEASURES: RESISTING NEUTRALITY, INEVITABILITY, AND INNOVATION NARRATIVES

As a first step in determining how AI should be regulated, we argue that lawmakers should acknowledge and resist the ideological narratives that motivate AI half measures, the first of which is the idea that AI systems are simply neutral conduits. A common misconception about technologies generally is that they are value-neutral, and this narrative has attached to AI as well. People often argue that AI systems can be used for pro-social or anti-social ends, but the technology itself isn't inherently good or bad. In other words, "there are no bad AI systems, only bad AI system users." This mistaken view leads to the common refrain that we should regulate uses of technology, not the technology itself.

This view of technologies is wrong. There is nothing value-neutral about any information technology, including AI systems. Values are deeply embedded into the design of technology human beings build. Every technology sends signals to people and makes a certain task easier or harder. Facial recognition technologies make people easier to surveil, which gives power to the watcher. Social media reduces the cost of speech and gaining attention, empowering those with the incentive to scale misinformation and disinformation efforts. Generative AI systems reduce the cost of countless tasks, reducing the value of certain labor and rewarding those who avoided the cost of the labor, even where that labor is a skill-developing Sophomore English essay. Lawmakers must take the design of AI systems seriously, looking to established theories of accountability like defective design and providing the means and instrumentalities of unfair and deceptive conduct.[80]

We also encourage lawmakers to resist the inevitability narrative of AI. Technologies like AI systems are not inevitable – they are intentionally designed and built by people, and people can prohibit them, they can regulate them, and they can shape their evolution into socially-beneficial

---

[79] *Id.*

[80] *See generally* WOODROW HARTZOG, PRIVACY'S BLUEPRINT, THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES (2018); Woodrow Hartzog, *Unfair and Deceptive Robots*, 74 Mar. L. Rev. 785 (2015).

tools as well. Lawmakers will make little progress until they accept that the toothpaste is never out of the tube when it comes to questioning and curtailing the design and deployment of AI systems for the betterment of society.[81]

The inevitability narrative often gets woven in with the ideology of "innovation" and cashed out as the necessity of "progress." But by itself, progress is an empty word. Progress for whom? Of what? At what cost? We cannot avoid making choices about what kind of future we want with AI. Similarly, a quasi-religious invocation of citing "tech progress" is not an answer either, much less a panacea. Like "progress," "innovation" is a buzz word that masks a thousand sins. Neil Richards has criticized tech company's repeated invocation of "innovation" as a canard.[82] The concept of innovation is selectively vague, meaning it can be whatever a tech company wants it to be, and to hear them tell it, innovation is always good and never bad.[83] It also has the strength of convenience—when advertising the latest product launch innovation seems like a supernatural force, but the moment regulation is proposed, innovation becomes easily "stifled," as fragile as a house of cards, toppled by the slightest regulation.[84]

Another branch of the Inevitability Narrative, which is parallel and, in many ways, contrary to the innovation and progress arguments, is that superintelligent AI poses an existential threat to humanity. Call this the macro existential question for AI. That existential threat, so it goes, is a

---

[81] Amba Kak & Sarah Myers West, *AI Now 2023 Landscape: Confronting Tech Power*, AI Now Institute (April 11, 2023), https://ainowinstitute.org/2023-landscape ("[T]here is nothing about artificial intelligence that is inevitable. Only once we stop seeing AI as synonymous with progress can we establish popular control over the trajectory of these technologies and meaningfully confront their serious social, economic, and political impacts.").

[82] *See* NEIL RICHARDS, WHY PRIVACY MATTERS 177-183 (2021) ("[H]ere's an experiment: take any sentence from a technology company about "innovation," and replace the word "innovation" with "magic" to see if the meaning of the sentence changes at all. In my own experience playing this game many times over the past decade, it almost never changes the meaning.").

[83] *Id.* ("The rhetorical construction of "innovation" by the tech sector slices off everything bad and leaves only the gleaming stainless steel of a technological utopia, one that is all Thomas More and no George Orwell.").

[84] *See* Neil Richards, Why Privacy Matters 177-183 (2021) ("[H]ere's an experiment: take any sentence from a technology company about "innovation," and replace the word "innovation" with "magic" to see if the meaning of the sentence changes at all. In my own experience playing this game many times over the past decade, it almost never changes the meaning.").

problem of such great magnitude that it therefore deserves regulatory and policy priority over current and near-term harms which stem from development and use of AI systems. Prominent AI ethicists have challenged this focus on the macro existential question for AI as distracting from immediate and more pressing AI harms.[85] Woodrow Hartzog has similarly challenged this narrative, arguing that "[i]f AI systems are used for existential harm, industry will be the engine—not some runway automation—and governments will be contributing or asleep at the wheel."[86]

Perhaps because of this consistent mislabeling of innovation and progress, it is easy to simply assume the rightful existence of AI systems and go straight to building guardrails so they can flourish. Sometimes this works well. For example, AI systems have the potential to help scientists solve vexing problems.[87] But it's dangerous to always assume the virtues of astonishingly powerful AI systems. For example, lawmakers in Maryland recently enacted a bill that will regulate police use of facial recognition technology.[88] Perhaps this law contains meaningful limits and safeguards on the use of such technology. But it nevertheless has the incidental effect of legitimizing this use of the technology. Professor Evan Selinger and Woodrow Hartzog have argued that some AI systems like face surveillance technologies are too dangerous to ever be safely deployed.[89] We recognize there is room for debate on this topic, but the point here is that when lawmakers go straight to putting up guardrails, they fail to ask the

---

[85] Devin Coldewey, *AI Ethicists Fire Back at 'AI Pause' Letter They Say 'Ignores the Actual Harms'*, TECH CRUNCH (Mar. 31, 2023), https://techcrunch.com/2023/03/31/ethicists-fire-back-at-ai-pause-letter-they-say-ignores-the-actual-harms (noting criticism from Timnit Gebru, Emily M. Bender, Angelina McMillan-Major and Margaret Mitchell that focus on "hypothetical risks" such as extermination via AI distracts from discussion of existing AI harms, such as "worker exploitation, data theft, [and] synthetic media that props up existing power structures and the further concentration of those power structures in fewer hands").

[86] Hartzog, *supra* note 13.

[87] Robert F. Service, *'The game has changed.' AI triumphs at protein folding*, SCIENCE (Dec. 4, 2020), https://www.science.org/doi/10.1126/science.370.6521.1144.

[88] S.B. 182, 2024 Md. Gen. Assembly, Reg. Sess (Md. 2024), https://mgaleg.maryland.gov/2024RS/bills/sb/sb0182E.pdf.

[89] *See* Evan Selinger & Woodrow Hartzog, *What Happens When Employers Can Read Your Facial Expressions?*, N.Y. TIMES (Oct. 17, 2019), https://www.nytimes.com/2019/10/17/opinion/facial-recognition-ban.html; Evan Selinger & Woodrow Hartzog, *The Inconsentability of Facial Surveillance*, 66 LOY. L. REV. 101 (2019).

existential question about whether particular AI systems should exist at all, and under what circumstances it should ever be developed or deployed. Call this the micro existential question, in contrast to the red herring macro existential question about AI-driven extinction. This principle also applies to bills that would regulate the use of automated decision-making technology for making certain consequential decisions such as employment, education enrollment, or access to essential goods. Even when tech companies initially resist the most dangerous tools like facial recognition, it seems unlikely that all participants in an unregulated industry can hold out forever when there is so much money on the table.[90]

Ignoring the true existential question about AI systems dooms us to a framework of half measures. We've already seen evidence of halfhearted approaches to limiting the rampant abuse of facial recognition technologies. Lawmakers and industry first welcomed these systems by demanding transparency, implementing ethical principles, mitigating bias, and requiring consent. Meanwhile companies big and small scraped every bit of biometric and personal data they could get on the open web under the dubious claim it was all "publicly available" and the flimsy pretext that collecting jaw-dropping amounts of data on every person who uses the Internet was necessary for training the system. AI half measures were no match for facial recognition systems, some of which were powered by the most astonishing and dangerous collection of data grabs I've ever heard of. These efforts have created untold implications for our privacy, our environment, and our democracy. Bright-line rules that prohibit certain kinds of design and deployments of AI systems and certain kinds of information collection and use can make sure that the development and deployment of AI is justified, and not a power grab that imposes massive external costs on society

III. SUBSTANTIVE INTERVENTIONS BEYOND HALF MEASURES

Industry leaders are quick to embrace AI half measures, touting the benefits of "responsible AI" and the self-regulation measures which they

---

[90] *See* Kashmir Hill, *The Technology Facebook and Google Didn't Dare Release*, The New York Times (Sept. 9, 2023), https://www.nytimes.com/2023/09/09/technology/google-facebook-facial-recognition.html.

believe will ensure that AI systems serve individuals and society.[91] 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 cadre" of good actors and conclude that no further action is needed.[92] 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,[93] or whether techniques like anomaly-detection "solve" the "black box" problem of AI systems,[94] 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.

The good news is that lawmakers, particularly state legislators, have started to get creative. Although not sufficient, privacy rules are necessary to properly regulate AI systems and state legislators have responded. [95] Senate Bill 2 in Connecticut, a proposed bill in 2024 which passed the state Senate but died under threat of a veto, included many substantive obligations such as impact assessments, risk management programs, and a duty of care which would have required developers and deployers of high-risk AI systems (those systems used to make certain consequential

---

[91] *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).

[92] *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").

[93] *See, e.g.*, *id.*; *see also* Solow-Niederman, *supra* note 50, 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).

[94] *See, e.g.*, Adebayo, *supra* note 91.

[95] For more on the relationship between privacy and artificial intelligence, see, e.g., Daniel J. Solove, Artificial Intelligence and Privacy, 77 Florida Law Review (forthcoming Jan 2025).

decisions) to avoid algorithmic discrimination.[96] A substantially similar bill was passed in Colorado on May 8th and sent to the Governor for signature.[97] A narrower but notable legislative innovation came in the proposed Minnesota Consumer Data Privacy Act, a bill developed over several legislative sessions by Minnesota state Representative Steve Elkins. That bill went beyond the standard individual rights which have come to typify U.S. state comprehensive privacy legislation by including a novel right for individuals to contest the result of profiling conducted in furtherance of decisions that produce legal or similarly significant effects.[98]

The prospect of omnibus AI legislation is daunting not only because it affects so many different areas of our life, but also because AI systems are so complex and powerful it can seem like trying to regulate magic.[99] But the broader risks and benefits of AI systems are not so new. AI systems bestow power. This power is used in all sorts of ways to benefit some and harm others. Some communities disproportionally benefit from that power, and other communities are marginalized and exploited by it. But there is good news: American law is no stranger to these power dynamics. As long as lawmakers keep inequalities and abuses of power and vulnerabilities to power at the center of their regulatory approach, they will be on the right track.

We recommend that since so many risks of AI systems come from within *relationships* where people are on the bad end of an information

---

[96] S.B. 2, Conn. Gen. Assembly, Reg. Sess. (Conn. 2024); Tatiana Rice, *Setting the Stage: Connecticut Senate Bill 2 Lays the Groundwork for Responsible AI in the States*, FUTURE OF PRIVACY (April 25, 2024), https://fpf.org/blog/setting-the-stage-connecticut-senate-bill-2-lays-the-groundwork-for-responsible-ai-in-the-states (describing the significance of S.B. 2 and arguing that it "would stand as the first piece of legislation in the United States governing the private-sector development and deployment of AI with comparable scale to the EU AI Act").

[97] S.B. 24-205, 74th Gen. Asm., Reg. Sess. (Colo. 2024), https://leg.colorado.gov/sites/default/files/documents/2024A/bills/2024a_205_rer.pdf.

[98] S.F. 4942, 93rd Leg., Reg. Sess. (Minn. 2024), https://www.revisor.mn.gov/bills/text.php?number=SF4942&version=latest&session=ls93&session_year=2024&session_number=0.

[99] *See* Efraín Foglia, Ferran Esteve, Lucía Lijtmaer, Luis Paadín, Óscar Marín, Miró Ramon & Mas Baucells, *Any sufficiently advanced technology is indistinguishable from magic*, CCCB LAB (Nov. 8, 2018), https://lab.cccb.org/en/arthur-c-clarke-any-sufficiently-advanced-technology-is-indistinguishable-from-magic/ (quoting Arthur Clarke's famous saying "Any sufficiently advanced technology is indistinguishable from magic."); Neil Richards, *Big data isn't magic*, ZOCALO PUBLIC SQUARE (Sept. 24, 2014), https://www.zocalopublicsquare.org/2014/09/24/will-we-have-any-privacy-after-the-big-data-revolution/ideas/up-for-discussion/#Neil+Richards.

asymmetry, lawmakers should implement broad, non-negotiable duties of loyalty, care, and confidentiality as part of any broad attempt to hold those who build and deploy AI systems accountable. Duties of loyalty protect against self- dealing, while related duties of care placed on relationships protect against dangerous behavior and the risks of harm. In other areas of the law, the extent of these duties is proportional to the vulnerability of the trusting parties.[100] The more exposed people are AI systems, the more loyalty, care, and confidentiality lawmakers should demand from those deploying the tools. These duties would provide a substantive prohibition on self-dealing and harm to limit AI systems in ways that half measures would not. And duties of loyalty, care, and confidentiality are tried-and-true tools that American law has used to mitigate power imbalances in relationships for literally hundreds of years.

A. Flexible Duties of Loyalty and Care

The conventional wisdom in tech policy that law cannot keep pace with technology is a gross misrepresentation.[101] While technology-specific rules might become outdates, more general 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.[102] As generative AI has

---

[100] *See, e.g.*, Hartzog & Richards, *supra* note 5; Jack M. Balkin, *The Fiduciary Model of Privacy*, 134 HARV. L. REV. F. 11 (2020) at 13–14.

[101] See, e.g., Josh Fairfield, Runaway Technology: Can Law Keep Up? (2021) (critiquing the assumptions built into the framing of the "pacing problem" of law lagging behind technology).

[102] 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*

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.[103] Basic principles of consumer protection law have proven extremely durable and responsive to changes in society, the economy, and technology. 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.[104] These kinds of flexible legal standards can enable accountability measures to be responsive to specific contexts and accommodate unknowns about downstream implementation.

1. Loyalty

Many of the problems of surveillance capitalism come down to the problem of self-dealing, where an organization exploits an advantage over a trusting party to its own benefit.[105] The lack of meaningful abilities to protect consumers under American privacy law has enabled such corporate opportunism and manipulation of consumers using human information, and this failure will only be exacerbated by the increased speed and

---

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.")

[103] 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.

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

[105] *See, e.g.*, JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM (2019); SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

efficiency of large language models to analyze and use this data. This problem is particularly serious in the context of AI systems and other technologies that promise to understand consumers so that they can better satisfy their needs and wants. Insufficiently constrained by privacy law and driven to maximize quarterly profits by corporate law, companies can deploy a potent cocktail of techniques derived from cognitive and behavioral science to "nudge" or otherwise influence the choices consumers make.[106] And history shows us that the companies that gather consumer data have not acted as benevolently as many had hoped.[107]

Misuse and self-enrichment through data gained in these power asymmetries ultimately costs consumers their time, money, attention, mental well-being, reputation, and significant life opportunities.[108] These costs include everything from their attention being broken via intrusive "notifications," to manipulation subtly shaping the way that consumers shop and vote, to the harms of engagement-driven social media.[109] "Personalization" of the companies' contacts and engagement strategies through the use of this personal data only magnifies these harms.[110] Such "personalization" can be finely calibrated to manipulate consumers into increasing engagement, regardless of any effect on consumers' mental wellbeing.[111] With every click and post, we are further exposed to the

---

[106] *See* NEIL RICHARDS, WHY PRIVACY MATTERS 39–50 (2022).

[107] *Id.*, *See generally,* RICHARD H. THALER & CASS. R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2008) (outlining how companies can promote pro-human outcomes if given the incentives necessary to induce those outcomes).

[108] *See* Neil Richards & Woodrow Hartzog, *Against Engagement,* B.U. L. Rev. (forthcoming 2024).

[109] *See generally*, Johann HARI, STOLEN FOCUS: WHY YOU CAN'T PAY ATTENTION – AND HOW TO THINK DEEPLY AGAIN (2022); Woodrow Hartzog & Neil Richards, *Legislating Data Loyalty*, 97 NOTRE DAME L. REV. REFLECTION 356 (2022)?

[110] This is precisely what happened in the Cambridge Analytica scandal, in which Facebook data was used to create finely calibrated psychological profiles of voters identifiedby their real names, suggesting which kinds of arguments would be most effective at gettingthem to act in the ways that the paying political advertisers wanted them to. *See* RICHARDS, *supra* note 9, at 25–26.

[111] These are the allegations Facebook whistleblower Frances Haugen presented under oath before lawmakers in the United States and around the world in 2021. *See, e.g.,* Billy Perrigo, *Inside Frances Haugen's Decision to Take on Facebook*, TIME (Nov. 22, 2021) https://time.com/6121931/frances-haugen-facebook-whistleblower-profile/ [https://perma.cc/L8QN-6GD5].

appetite, carelessness, and influence of those developing and deploying AI systems.

AI systems will never work for the benefit of all unless society can reliably trust those designing and deploying them. Right now, the trust people are giving these companies is a blind trust that is regularly betrayed. What is needed are rules that make companies deploying AI systems trust-*worthy*. This is where duties of loyalty, care, and confidentiality come in. The core feature of a duty of loyalty is that it creates a substantive duty prohibiting self-dealing at the expense of a trusting party.[112] Relational duties, such as a duty of loyalty, offer distinct advantages for lawmakers looking to address privacy across multiple disparate actors and methods of data consumption.

First, relational duties are sensitive to power disparities within information relationships. Second, relational duties help to mitigate the issues with overwhelming corporate disclosures and requests for consent. Relational duties allow lawmakers to move beyond ineffective consent frameworks while preserving meaningful choices for people. These duties allow trusting parties to enter information relationships without accepting the risks of whatever harmful data practices and consequences lurk in the fine print, the business model, or the technology.[113] They can also allow a broader range of potential choices because under a duty of loyalty, people are protected regardless of what they choose.[114] Relationships open the possibility of more robust enforcement rules because they are voluntarily entered into and because they are more consistent with free expression principles. This is why relational rules have long been recognized in American law.[115]

A substantive duty of data loyalty could revolutionize American privacy law. As we have argued in previous articles, comments, and

---

[112] *See generally*, U.S. Sen. Comm. on Health Educ. Lab. & Pensions, Comment Letter on Improving Americans' Health Data Privacy (Sept. 28, 2023), (on file with The Cordell Institute for Policy in Medicine & Law); Hartzog & Richards, *supra* note 5.

[113] *Id.*

[114] For an extended critique of consent-based models for data processing, *see* Hartzog & Richards, *supra* note 5; *see also* Richards & Hartzog, *supra* note 11, and HARI, *supra* note 12.

[115] HARI *Supra* note 12.

testimony,[116] we believe that creating a broad duty of data loyalty offers three important advantages that other approaches do not. First, a duty of loyalty is substantially more able than a traditional data protection approach to address the novel problems created by the explosion of "big data" processing and analytics.[117] These include algorithmic discrimination, manipulation, oppression, and shaming that are caused by the ubiquity of modern technology platforms.

Second, loyalty helps solve privacy law's harm problem in a way that is consistent with the direction of current Supreme Court doctrine. The exploitation of a relationship against a trusting party's interests, such as in a case of conflict of interest, can be a legally-cognizable concrete harm even if no other tangible harm manifests.[118] This is significant because American plaintiffs in privacy and data breach lawsuits have struggled to articulate harm that courts will recognize, particularly as the federal courts have tightened the rules for what constitutes a recognizable harm.[119] By contrast, because our common law duties of loyalty are literally older than the United States itself, they offer a tried and tested mechanism to resolve the power imbalances in relationships like those between doctors and patients and platforms and consumers.

---

[116] *See e.g.,* Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. 961; U.S. Sen. Comm. on Health Educ. Lab. & Pensions, Comment Letter on Improving Americans' Health Data Privacy (Sept. 28, 2023), (on file with The Cordell Institute for Policy in Medicine & Law); Fed. Trade Comm'n, Comments of the Cordell Institute on the Prevalence of Commercial Surveillance and Data Security Practices that Harm Consumers (Nov. 22, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4284020; U.S. Senate Comm. on the Judiciary, Subcomm. on Priv. Tech. and the L., Testimony of Woodrow Hartzog on "Legislating of Artificial Intelligence" (Sept. 12, 2023), https://www.judiciary.senate.gov/imo/media/doc/2023-09-12_pm_-_testimony_-_hartzog.pdf; Nat'l Telecomm. and Info. Admin., Comments of the Cordell Institute on AI Accountability (June 12, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4477426.

[117] *See generally* Woodrow Hartzog & Neil Richards, *Trusting Big Data Research*, 66 DEPAUL L. REV. 579 (2017)

[118] *See* TAMAR FRANKEL, FIDUCIARY LAW 107–08 (2011) ("The duty of loyalty supports the main purpose of fiduciary law: to prohibit fiduciaries from misappropriating or misusing entrusted property or power. Thus, the duty of loyalty is manifested by important preventative rules. Such rules prohibit actions even though they are not necessarily injurious to entrustors."); see also *Spokeo v. Robbins* 136 S.Ct. 1540 (2016).

[119] See, e.g., *TransUnion LLC v. Ramirez,* 141 S.Ct. 2190 (2021); *Spokeo, Inc. v. Robins,* 136 S.Ct. 1540 (2016).

A third benefit of a loyalty-based approach to privacy law is that loyalty duties have a long and established development in our law, most famously in the law of fiduciaries. A duty of data loyalty could draw heavily from this tradition and its proven ability to protect against the power imbalances in relationships in a fair, principled, and meaningful way. (We note that the professional ethics of both lawyers and doctors already require that they be loyal to their clients and patients; perhaps those of data scientists should as well.)[120]

Loyalty, care, and confidentiality are not just foundational concepts in American law, they are also deeply intuitive. Lawmakers should not underestimate loyalty's rhetorical potential. A rallying cry requiring companies to "act in our best interests" could motivate American privacy reform in the way that "the right to be let alone" did at the turn of the twentieth century. Technocratic terms like "data minimization" and "legitimate interests of the data controller" do little for public imagination or comprehension. By contrast, loyalty is clear, it is easy to understand, and it is potentially robust enough to counterbalance spurious industry claims about the importance of "innovation" or the idea that commercial data processing implicates significant First Amendment issues. GDPR-style ideas like requiring companies to undergo data protection impact assessments can feel wonky and feeble, but every person in America likely knows how it feels to be betrayed.

There is also a roadmap for lawmakers looking to impose duties of loyalty on the powerful. Fiduciary law scholars have identified a tried and tested two-step process that lawmakers use to implement loyalty obligations in such a fair and just way.[121] Lawmakers first articulate a primary, general duty of loyalty—one that can be relatively permissive but acts as a residual backstop against betrayal. Second, courts and lawmakers go about the task of creating and refining what have been referred to as

---

[120] Richards & Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. 961, 968 (2021).

[121] *See e.g.* Robert H. Sitkoff, *Other Fiduciary Duties: Implementing Loyalty and Care* in THE OXFORD HANDBOOK OF FIDUCIARY LAW 419 (Evan J. Criddle et al., 2019). ("The duties of loyalty and care, which we might call the primary fiduciary duties, are typically structured as broad, open-ended standards that speak generally. By contrast, the other fiduciary duties, which we might call the subsidiary or implementing fiduciary duties, are typically structured as rules or at least as more specific standards that speak with greater specificity.").

"subsidiary" duties that are more specific and sensitive to context. These subsidiary duties target the most opportunistic contexts for self-dealing and typically result in a mix of overlapping open-ended rules, maxims, more specific standards, and context-specific rules.

Thus, we propose that a duty of data loyalty should be implemented on two levels through what we have called the "loyalty two-step."[122] The first level is a broad and general "catch all" prohibition on substantial conflicts with the trusting party's best interests. This would prevent the most egregious forms of disloyalty across the board, and it would also serve to orient the company's incentives generally against betrayal rather than micromanaging specific instances. It would also supply a backstop against novel or innovative forms of betrayal that allows the law to evolve for new circumstances.

The second level subsidiary duty of loyalty rules should be more specific and, where necessary, restrictive. This would involve the articulation of specific and substantive rules targeting particular contexts and actions that provide clearer rules than the general duty and would leave less wiggle room to ensure accountability. This clarity will keep the frameworks from becoming watered down. In the health care context, for example, bright-line rules should be more restrictive where companies are using personal health data for marketing or persuasion, or where they are collecting location data, but more permissive where personal health data is being used for biomedical research in the public interest. Through this layered approach, a duty of data loyalty could provide both general applicability as well as sensitivity to specific contexts.[123]

There is already bipartisan support for a duty of loyalty, including the proposed American Data Privacy and Protection Act (ADPPA).[124] However, the best starting point for statutory language is the proposed bipartisan Digital Consumer Protection Commission Act of 2023 (DCPCA)

---

[122] *See* Hartzog & Richards, *supra* note ^.

[123] *Id.*

[124] *See* Woodrow Hartzog & Neil Richards, *We're So Close to Getting Data Loyalty Right*, INTERNATIONAL ASSOCIATION OF PRIVACY PROFESSIONALS (June 14, 2022), https://iapp.org/news/a/were-so-close-to-getting-data-loyalty-right/.

for online platform regulation.[125] The relevant language appears in Section 2411:

> "**SEC. 2411. DUTY OF LOYALTY.**
>
> A covered entity may not process personal data or design information technologies in a way that substantially conflicts with the best interests of a person with respect to—
>
> (1) the experience of the person when using a platform owned or controlled by the covered entity;
>
> or
>
> (2) the personal data of the person."[126]

We believe that a duty of loyalty provides the strongest, and most comprehensive protections against the misuse of consumer data by the creators and deployers of AI systems. This duty, implemented through the loyalty two-step described above could address the potentially hungry, leaky, sneaky, and exclusory effects of the unfettered proliferation of AI systems.

2. Care

While duties of loyalty are vital for preventing exploitation and abuse, duties of care are essential for preventing harm, regardless of whether companies are engaged in self-dealing. Duties of care are key to help set baseline safety standards that draw from custom and normative notions of reasonable behavior for companies designing systems that, by their nature, create (and/or reduce) varying risks of harm for people. Furthermore, duties of care have proven resilient and flexible in response to technological change throughout the decades.

Lawmakers are already experimenting with a duty of care for AI. In the 2024 state legislative sessions, one of the most significant AI regulatory bills introduced was Connecticut Senate Bill 2.[127] In addition to becoming

---

[125] Digital Consumer Protection Commission Act of 2023, S. 2597, 118th Cong. (2023).

[126] *Id.* at §2411.

[127] For an overview of the substantive requirements of SB 2, see Tatiana Rice, *Connecticut Senate Bill 2 Two-Pager Cheat Sheet*, FPF (April 25, 2024), https://fpf.org/wp-content/uploads/2024/04/FPF-FINAL-CT-SB-2-Two-Pager-4-25-24.pdf.

a model bill for other states, one of the most notable features of SB 2 was its inclusion of a duty of care. Under the version of the bill that passed the Senate in April 2024, deployers of high-risk AI systems would be required "to protect consumers from any known or reasonably foreseeable risks of algorithmic discrimination."[128] For developers, the duty was limited to protecting from such risks "arising from the intended and contracted uses of such high-risk artificial intelligence system."[129]

The bill also included a rebuttable presumption that a developer or deployer used reasonable care if they complied with their respective requirements under the relevant section. For developers, these requirements included: making certain information regarding the high-risk AI system available to deployers (i.e., intended use, types of data used to train the system, how the system was evaluated for performance, *et cetera*); providing information and documentation necessary for deployers to conduct impact assessments; and making available to the public information certain information about the high-risk AI systems developed.[130] For deployers, these requirements included: maintaining a risk management policy and program in accordance with standards under the bill; conducting impact assessments; annually reviewing deployed high-risk AI systems to ensure they are not causing algorithmic discrimination; providing notice to the consumer that a high-risk AI system is in use; if an adverse consequential decision was reached, notice and opportunity to appeal with human review where technically feasible; and providing certain public information regarding the high-risk AI systems in use.[131]

A prior draft of the bill from when it was in committee would have included a much broader duty of care for developers of generative AI systems, requiring reasonable care to protect individuals from known or reasonably foreseeable risks of: unfair or deceptive trade practices; unlawful disparate impact; redressable emotional, financial, mental, physical or reputational injury; redressable intrusion upon solitude or

---

[128] S.B. 2, Conn. Gen. Assembly, Reg. Sess., § 3 (Conn. 2024), https://www.cga.ct.gov/2024/lcoamd/pdf/2024LCO04463-R00-AMD.pdf.

[129] *Id.* § 2.

[130] *Id.* § 2.

[131] *Id.* § 3.

seclusion of private affairs or concerns; or to intellectual property rights.[132] Although this language was not included in the version of the bill which passed the Connecticut Senate, that it was included in a public committee draft shows that US lawmakers increasingly are aware of the broad array of risks to individuals that flow from AI systems and see a flexible duty of care as a potential means to prevent those harms.

B. Specific Rules

Fostering trust and accountability in the AI ecosystem will require supplementing flexible duties with specific rules aimed at specific practices as well as effectuating more structural change. We specifically recommend design rules, outright prohibitions, structural support, and private causes of action for more effective enforcement.

Woven together as a comprehensive regulatory fabric, these duties, rules, and commitments can invigorate and strengthen procedural tools such as audits and certifications, to the benefit of people both individually and as a group. These duties could be crafted not only to protect our privacy, but also implemented in frameworks designed to protect our health and wellbeing, environment, and workplace.

1. Design Rules and Secondary Liability

On a more substantive level, emboldened consumer protection rules, such as prohibitions on unfair, deceptive, and abusive acts or practices, can also help protect our trust. Lawmakers could implement substantive prohibitions like robust data minimization rules, limits on particular uses of data, and prohibitions on abusive design practices like dark patterns and predatory algorithms.[133]

---

[132] Joint Gen. L. Comm., S.B. 2, Conn. Gen. Assembly, Reg. Sess. (Conn. 2024), https://www.cga.ct.gov/2024/TOB/S/PDF/2024SB-00002-R01-SB.PDF (early version of SB 2 drafted by the Joint Committee on General Law on February 21, 2024).

[133] Computer science and policy scholarship demonstrates dark patterns' pervasiveness in consumers' digital lives. *See generally* Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt & Austin L. Toombs, *The Dark (Patterns) Side of UX Design*, EXTENDED ABSTRACTS OF THE 2018 CHI CONFERENCE ON HUMAN FACTORS IN COMPUTING SYSTEMS (Apr. 20, 2018), https://pure.psu.edu/en/publications/the-dark-patterns-side-of-ux-design; Arunesh Mathur, Gunes Acar, Michael J Friedman, Elena Lucherini, Jonathan Mayer, Marshini Chetty & Arvind Narayanan, *Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites*, 3 PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER

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,[134] 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.

Policymakers should also 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.[135] In a complaint filed against a geolocation data broker, for example, the FTC alleged that the sale of precise geolocation data "could enable third parties

---

INTERACTION 1-32 (Nov. 7, 2019), https://dl.acm.org/doi/10.1145/3359183; Johanna Gunawan, David Choffnes, Woodrow Hartzog & Christo Wilson, *A Comparative Study of Dark Patterns Across Mobile and Web Modalities*, 5 PROCEEDINGS OF THE ACM ON HUMAN-COMPUTER INTERACTION 1-29 (Oct. 18, 2021), https://dl.acm.org/doi/10.1145/3479521; Jamie Luguri and Lior Jacob Strahilevitz. 2021. Shining a Light on Dark Patterns. Journal of Legal Analysis 13, 1 (March 2021), 43–109; Kentrell Owens, Johanna Gunawan, David Choffnes, Pardis Emami-Naeini, Tadayoshi Kohno & Franziska Roesner. *Exploring Deceptive Design Patterns in Voice Interfaces*, EUROUSEC '22: PROCEEDINGS OF THE 2022 EUROPEAN SYMPOSIUM ON USABLE SECURITY (Sept. 29, 2022), https://dl.acm.org/doi/10.1145/3549015.3554213.

[134] 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.

[135] *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.").

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.'"[136] Holding companies liable for their conduct which enables secondary harms will be a critical tool in creating trustworthy and accountable AI systems.

2. Bright-line Prohibitions

We join the groups Accountable Tech, AI Now, and the Electronic Privacy Information Center in their call for bright-line rules.[137] As part of the implementation of duties of loyalty, care and confidentiality, lawmakers should prohibit unacceptable AI practices like emotion recognition, unconstrained facial recognition, predictive policing, remote biometric identification in social spaces, social scoring, and fully automated hiring and firing. They should prohibit most secondary uses and third-party disclosure of personal data and while also requiring protections against third party access, including data-scraping.

3. Structural and *Ex Ante* Strategies

To begin, if regulators want to avoid their frameworks being co-opted and diluted by industry, they should consider more structural and systemic approaches and lenses. Julie Cohen has recently argued in favor of moving online governance to infrastructure to counteract industry's work to bend infrastructure of the public sphere to their own self-interested and profit-maximizing purposes.[138] This will entail "some rethinking of traditional assumptions equating structural control with censorship and privatization

---

[136] 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).

[137] *Zero Trust AI Governance*, AI NOW INSTITUTE (Aug. 2023), https://ainowinstitute.org/wp-content/uploads/2023/08/Zero-Trust-AI-Governance.pdf.

[138] Julie Cohen, Infrastructuring the Digital Public Sphere, 25 Yale J.L. & Tech. Special Issue 1 (2023) ("current patterns of online communication are not inevitable features of the digital public sphere's natural evolution but rather are the result of infrastructuring work undertaken for particular, self-interested purposes. Patterns of online communication flows now engineered systemically for maximum volatility and virality might be engineered differently, and free speech law for the digital public sphere might be reenvisioned as permitting—or even requiring—public governance mandates that attempt to restore conditions of flow more compatible with the survival and healthy functioning of democratic institutions.")

with counterpower. It also necessitates some careful rethinking of regulatory targets and methods."[139] Infrastructure approaches are similar to design approaches, but Cohen argues that while design-thinking focuses on affordances and how people perceive and react within environments, infrastructure thinking "probes downward and outward to consider the underlying, habituated arrangements through which activities of exchange and the social orderings they produce are enabled and shaped at scale."[140] Cohen argues that "the quest for fair choice architectures has a way of rendering underlying arrangements for data harvesting and real-time, data-driven patterning invisible; infrastructure thinking aims to expose those arrangements and consider what they ask us to take for granted."[141] Infrastructure thinking is a direct antidote to inevitability narratives, because it encourages lawmakers to question the existence and purpose of things like algorithmic optimization, software development kits, and platform advertising dashboards.[142]

Other structural approaches could involve changes within government itself. For example, a new federal agency could coordinate these protective efforts across various legal frameworks. For example, the bipartisan AI legislation framework released by U.S. Senators Blumenthal (D-CT) and Hawley (R-MO) included establishing an independent oversight body which would, in part, cooperate with other enforcers.[143] Lawmakers should fund existing agencies and empower them to hire technologists and better enforce regulations within their existing expertise, but there is a potential role to play for an expert "meta" agency to provide expertise and coordination amongst government actors.

Licensing is another way to structurally ensure that an institution must respect human values and prove themselves worthy of our trust as a prerequisite to market entry. This *ex ante* approach could be a powerful correction to business models that encourage harmful behavior. It would require a sound basis for processing data and deploying technologies. A

---

[139] *Id.* at 29.
[140] *Id.* at 17.
[141] *Id.*
[142] *Id.*
[143] Press Release, *Blumenthal & Hawley Announce Bipartisan Framework on Artificial Intelligence Legislation* (Sept. 8, 2023), https://www.blumenthal.senate.gov/newsroom/press/release/blumenthal-and-hawley-announce-bipartisan-framework-on-artificial-intelligence-legislation.

justification-first approach combined with substantive duties and strong liability rules would also flip the presumption that burdens society with the risk of dangerous systems by requiring companies to justify their systems by proving they will not harm us.[144]

Other structural rules would also help hold developers and deployers accountable while simultaneously benefitting all technology law and regulatory efforts. Lawmakers must protect researchers and whistleblowers. They should refine the Computer Fraud and Abuse Act and create an avenue for researchers to discover abuses of and within AI systems while preserving the trust of people exposed to those systems. They should expand public policy exceptions to NDAs for whistleblowers to report those abuses as in California and Washington's Silenced No More Acts. Governments should invest in their own expertise by ensuring a broad range of methodological experts are staffed in every office that deals with AI systems. Give the Federal Trade Commission, an agency already deeply invested in protecting the public from the abuses of AI systems, more staff and more funding and modify the prohibition on unfair and deceptive trade practices to include a prohibition on abusive trade practices (as at least one state has already done) and remove the cost-benefit requirement in Section 5 that fails to properly elevate human values and wellbeing.

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.[145] 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.

---

[144] *See generally* Gianclaudio Malgieri & Frank Pasquale, *From Transparency to Justification: Toward Ex Ante Accountability for AI*, 712 Brook. L. Sch. Legal Stud. (2022).

[145] *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.

4. Private Right of Action and Hybrid Enforcement

In our body of work advocating for imposing a duty of loyalty in privacy law, we have previously the necessity of including a private right of action in any comprehensive consumer data privacy legislation.[146] Lauren Henry Scholz's insights on the importance of a hybrid enforcement regime are as true for artificial intelligence as they are for privacy.[147]

Notwithstanding the merits of private litigation, it is equally important to adequately fund regulators to enforce privacy and consumer protection rules in the context of AI. The FTC has long shown its ability to obtain meaningful, substantive consent decrees from companies for violations of the FTC Act's prohibition on unfair and deceptive trade practices. The FCC has similarly flexed its muscles to obtain significant monetary fines and injunctive relief in its area of expertise. In April 2024, the FCC announced that it had fined several wireless carriers nearly $200 million for "for illegally sharing access to customers' location information without consent and without taking reasonable measures to protect that information against unauthorized disclosure."[148] This kind of institutional knowledge and structural leverage is invaluable for obtaining meaningful relief for individuals harmed or exposed to unreasonable risk by AI systems. Although AI entrepreneurs might recommend the establishment of a standalone AI agency,[149] it is clear that policymakers can choose to fund existing regulators now to enforce new and existing law against AI developers and deployers.

---

[146] Neil Richards, Woodrow Hartzog & Jordan Francis, *A Concrete Proposal for Data Loyalty*, 37 HARV. J. LAW & TECH. 1335, 1360–61 (2024) ("A robust private right of action is vital for ensuring optimal levels of enforcement.").

[147] Lauren Henry Scholz, *Private Rights of Action in Privacy Law*, 63 Wm. & Mary L. Rev. 1639, 1644–45 (2022) ("Private rights of action have two important benefits for privacy regulation. First, private enforcement marshals the resources of the private sector to fund and provide information in dealing with this ubiquitous issue . . . . Second, private rights of action have expressive value that cannot be achieved through public regulation in the area of privacy.").

[148] Press Release, Fed. Commc'ns Comm'n, FCC Fines AT&T, Sprint, T-Mobile, and Verizon Nearly $200 Million for Illegally Sharing Access to Customers' Location Data (Apr. 29, 2024), https://docs.fcc.gov/public/attachments/DOC-402213A1.pdf.

[149] Cat Zakrzewski, Cristiano Lima-Strong & Will Oremus, *CEO Behind Chatgpt Warns Congress AI Could Cause 'Harm to the World'*, WA. POST (May 16, 2023), https://www.washingtonpost.com/technology/2023/05/16/sam-altman-open-ai-congress-hearing (noting that in his testimony before the U.S. Congress, OpenAI CEO called for the establishment of a new federal agency tasked with regulating AI).

***

Woven together as a comprehensive regulatory fabric, these duties, rules, and commitments can invigorate and strengthen procedural tools such as audits and certifications, to the benefit of people both individually and as a group. These duties could be crafted not only to protect our privacy, but also implemented in frameworks designed to protect our health and wellbeing, environment, and workplace.

CONCLUSION

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 discussions of AI safety, accountability, ethics, and fairness: "The concept of progress acts as a protective mechanism to shield us from the terrors of the future."[150] 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 this article, we have argued that the bulk of industry-led AI policy approaches over the past several years, such as encouraging transparency, mitigating bias, promoting ethical principles, and giving people choice, are vital, but *they are not enough*, whether individually or collectively. In the end, they will not fully protect society, at the risk of giving the impression that our rules are sufficient and that lawmakers have done enough. In that sense these measures are best thought of as "AI half measures."

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."[151] 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

---

[150] FRANK HERBERT, DUNE 410 (Penguin Books 2016) (1965).
[151] Hartzog & Richards, *supra* note 104, at 990.

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*. Instead, 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. But we must be cautious, and we have good reason to be skeptical. AI systems also pose huge risks to almost every important aspect of our lives. They are already being used to amass power, harm vulnerable people, and erode social and political institutions. As AI systems kickstart a new phase of information systems revolution, we must avoid the mistakes of the past and proactively approach the difficult 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.