PREPARED TESTIMONY AND STATEMENT FOR THE RECORD
OF

WOODROW HARTZOG
PROFESSOR OF LAW, BOSTON UNIVERSITY SCHOOL OF LAW
FELLOW, CORDELL INSTITUTE FOR POLICY IN MEDICINE & LAW, WASHINGTON UNIVERSITY IN ST. LOUIS

on behalf of himself and based on research with scholars from the Cordell Institute for Policy in Medicine & Law, Washington University in St. Lous:

NEIL RICHARDS
KOCH DISTINGUISHED PROFESSOR IN LAW AT WASHINGTON UNIVERSITY SCHOOL OF LAW AND CO-DIRECTOR OF THE CORDELL INSTITUTE

RYAN DURRIE
ASSOCIATE DIRECTOR FOR POLICY, CORDELL INSTITUTE

September 12, 2023

HEARING ON

"OVERSIGHT OF A.I.: LEGISLATING ON ARTIFICIAL INTELLIGENCE"

BEFORE THE

U.S. SENATE COMMITTEE ON THE JUDICIARY
SUBCOMMITTEE ON PRIVACY, TECHNOLOGY, AND THE LAW

I. INTRODUCTION

Chair Blumenthal, Ranking Member Hawley, and Members of the Committee, thank you for inviting me to appear before you and provide testimony on this important issue. My name is Woodrow Hartzog and I am a Professor of Law at Boston University School of Law and a Non-Resident Fellow at The Cordell Institute for Policy in Medicine & Law at Washington University in St. Louis. I have written extensively on privacy and law and technology issues, including how the unregulated amassing of personal data has fueled high-risk uses of artificial intelligence or "AI". My comments today will address what I've learned from this research. I make these comments in my personal, academic capacity. My remarks are based on research I have conducted in conjunction with scholars from the Cordell Institute for Policy in Medicine & Law at Washington University in St. Louis.[1]

I applaud the Committee for taking the initiative and advancing the conversation about legislating artificial intelligence. People and institutions are using AI systems to reshape how we work, play, socialize, and civically engage. These systems have proven useful, but they also impose an enormous cost on our wellbeing, our environment, and our democracy. So, we must make sure they are worth it for reasons beyond making things cheaper.

In my testimony today, I'll discuss how 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."

I'd like to make one simple point in my testimony today: ***To bring AI within the rule of law, lawmakers must go beyond half measures to ensure that AI systems and the actors that deploy them are worthy of our trust***.[2] To do that, lawmakers must do three things.

[1] The Cordell Institute works to be an internationally recognized voice and resource in ethical health and human information policy, promoting the compelling societal good of advancing research and discovery into human wellness and disease while protecting the privacy and autonomy of patients, consumers, and citizens. We also thank Agnish Chakraburtty for his assistance with these comments.

[2] My Cordell Institute colleague Professor Neil Richards and I have argued that trust and relational vulnerability are the critical lenses through which to view issues of privacy, data protection, and civil rights in the digital age. We believe that issues of trust and vulnerability are equally critical for the design and implementation of AI systems that affect our minds, our health, our life opportunities, our finances, our labor, our markets, our environment, and even our democracy itself. *See, e.g.*, Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. 961; Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431(2016); Neil Richards & Woodrow Hartzog, *Privacy's Trust Gap: A Review*, 126 YALE L.J. 1180 (2017); Woodrow Hartzog & Neil Richards, *Trusting Big Data Research*, 66 DEPAUL L. REV. 579 (2017); Neil Richards & Woodrow Hartzog, *A Relational Turn for Data Protection?,* 6 EUR. DATA PROT. L. REV. 492 (2020); Woodrow Hartzog & Neil Richards, *The Surprising Virtues of Data Loyalty*, 71 EMORY L.J. 985 (2021); Woodrow Hartzog & Neil Richards, *Legislating Data Loyalty*, 97 NOTRE DAME L. REV. REFLECTION 356 (2022).

*First, they must accept that AI systems are not neutral***.** This includes moving swiftly in holding developers of AI systems accountable for their design choices.

Second, they must focus on substantive interventions that limit abuses of power. Such approaches include imposing duties of loyalty, care, and confidentiality and *ex ante* approaches requiring a sound basis for processing data and deploying technologies.

Third, they must resist the narrative that AI systems are inevitable by 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.

II. AI HALF-MEASURES ARE IMPORTANT, BUT INADEQUATE

In work with scholars at the Cordell Institute, we have argued that establishing trustworthy and accountable AI systems requires avoiding or at least going beyond "AI half-measures."[3] We have conceived of these half-measures as necessary but insufficient approaches that have a limited efficacy and in isolation may give lawmakers a false sense of security that they've done enough. They also risk tricking people into thinking new AI systems give little cause for worry.

Half measures often take the form of post-deployment controls, audits, assessments, certifications, and similar procedural compliance requirements.[4] 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.[5] 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.

The AI landscape is at a crossroads and now is the time to act. The harms of AI are real, significant, and becoming both entrenched and normalized by the day.[6] If we do not

[3] Neil M. Richards, Woodrow Hartzog & Jordan Francis, *Comments of the Cordell Institute on AI Accountability* (June 12, 2023), https://www.regulations.gov/comment/NTIA-2023-0005-1291 (Comment ID: NTIA-2023-0005-1291); *see also* 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), https://www.regulations.gov/comment/FTC-2022-0053-1071 (Comment ID: FTC-2022-0053-1071).

[4] *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").

[5] *See generally*, ARI EZRA WALDMAN, INDUSTRY UNBOUND: THE INSIDE STORY OF PRIVACY, DATA, AND CORPORATE POWER (2021).

[6] *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. (forthcoming), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4384541.

impose rules to limit abuses of power, we risk eroding our civil liberties, our civil rights, and our democracy itself. The widespread adoption of AI systems implicates many of the rights and values we hold most dear: due process, freedom of expression, anti-discrimination, voting, free and fair elections, privacy, identity formation, the integrity of meaningful and intimate social relationships, and opportunities for safe, healthy, and fulfilling work.[7]

If lawmakers hope to foster trust and accountability in AI systems, they must avoid the seductive appeal of relying solely on "AI half-measures"—regulatory tools and mechanisms like transparency requirements, bias audits, ethical principles, and other procedural requirements that are necessary but not sufficient for true accountability. When implemented as standalone protections rather than as components of broader governance strategies, AI half-measures provide merely 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, a commitment solely to AI half-measures reveals itself as pernicious—offering the illusion of protection while enabling the festering of harms and other social costs. This might make AI half-measures appealing from an industry perspective but it definitely makes them dangerous for society. In my testimony today, I will identify four such AI half-measures which are necessary but not sufficient for true accountability: transparency, bias mitigation, ethics, and individualistic control.

A. TRANSPARENCY DOES NOT AUTOMATICALLY MAKE THINGS RIGHT WHEN THINGS GO WRONG

At best, AI transparency can only be a first step, and not an end in itself. Certainly, AI systems themselves are astonishingly opaque at a scale "almost beyond human imagining."[8] As decision making concerning ordinary people increasingly relies on AI systems, we have a right to know the basis upon which those decisions are being made. People subject to AI system decision-making deserve "technological due process" that provides meaningful notice and transparency.[9] But transparency isn't always helpful to people without power. Even if you are aware a system is in use you might not be able to

[7] *See generally*, JULIE COHEN, CONFIGURING THE NETWORKED SELF (2012); DANIELLE CITRON, THE FIGHT FOR PRIVACY: PROTECTING DIGNITY, IDENTITY, AND LOVE IN THE DIGITAL AGE (2022); DANIEL SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE (2006); DANIEL SOLOVE, THE FUTURE OF REPUTATION: GOSSIP, RUMOR, AND PRIVACY ON THE INTERNET (2008); 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); KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE (2021); FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2016); FRANK PASQUALE, NEW LAWS OF ROBOTICS: DEFENDING HUMAN EXPERTISE IN THE AGE OF AI (2020); SAFIA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018); IFEOMA AJUNWA, THE QUANTIFIED WORKER: LAW AND TECHNOLOGY IN THE MODERN WORKPLACE (2023).

[8] *See generally*, Kate Crawford & Vladan Joler, *Anatomy of an AI System*, MOMA (2018) https://anatomyof.ai; *accord* KATE CRAWFORD, ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE (2021).

[9] *See generally*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249 (2008).

do much about it. It is critical, therefore, that lawmakers also have some level of insight into the design and implementation of these systems.

But looking into a system does not necessarily lead to knowing about it.[10] 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.

Transparency is a pleasant-sounding word, but in practice it can mean very different things to different actors in the AI conversation. The importance of transparency in AI accountability, however, 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 look into these systems; we must understand their various parts and change them when they threaten harm and betrayal.

B. BIAS MITIGATION PROPOSALS ALONE ARE NOT ENOUGH.

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.[11] To mitigate bias, lawmakers, regulators, and those in industry call

[10] *See* Mike Ananny & Kate Crawford, *Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability*, 20 NEW MEDIA & SOCIETY 973, 977-982 (2018). Scholars Mike Annay and Kate Crawford 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. 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. 2.Full transparency can be harmful, especially to vulnerable individuals. 3.Transparency can intentionally occlude by making so much information available that it conceals more damaging information. 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." 5.Transparency burdens individuals by forcing them to "seek out information about a system, to interpret that information, and determine its significance." Similarly, the transparency ideal also presumes that different systems can easily be compared, allowing individuals to assess and choose between alternative options. 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. 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. 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. 9.Technical limitations—resulting from the scale and speed of a system's design—can make a system inscrutable, even to its creators. This problem is especially challenging in the context of machine learning AI systems such as deep learning. 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."

[11] *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

for more care in the development of AI systems.[12] I agree. AI systems will remain dangerously and possibly fatally flawed so long as they reflect harmful societal discriminatory practices. Bias mitigation is undoubtedly worthy of attention, resources, and regulation. 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 highly contested. 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. After all, "the very choice of a mathematical definition of 'fairness' is a political one."[13] Lawmakers must make that difficult decision and give the concept of fairness a clear and firm moral anchor. Leaving the definition of bias to the collective wisdom of self-interested individuals will lead to splintered and self-serving definitions that are likely to enable more bad practices than they prevent. Not to put too fine a point on it, but ***AI self-regulation will be doomed to fail***. Lawmakers should treat disparate impacts of AI systems as an issue of civil rights while also enforcing existing privacy and consumer protection laws.[14]

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

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).

[12] *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))).

[13] Alicia Solow-Niederman, *Information Privacy and the Inference Economy*, 117 NW. U. L. REV. 357, 420 (2022) (*citing* Arvind Narayanan, *Tutorial: 21 Fairness Definitions and Their Politics*, YOUTUBE (Mar. 1, 2018), https://www.youtube.com/watch?v=jIXIuYdnyyk).

[14] *See generally* Solon Barocas and Andrew Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671 (2016); Andrew Selbst and Solon Barocas, *Unfair Artificial Intelligence: How FTC Intervention Can Overcome the Limitations of Discrimination Law*, 171 U. PENN. L. REV. 1023 (2023).

system is less dangerous. But accurate AI systems are, if anything, more dangerous.[15] 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.[16] 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.

C. ETHICAL PRINCIPLES ARE NOT ENOUGH WITHOUT ENFORCEMENT

Ethics should certainly be front of mind for companies that design or deploy AI systems. 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.[17] 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.[18] 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 companies continuing to develop dubious AI systems that violate their own ethics principles. [19]

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

[15] *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.

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

[17] 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/.

[18] 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.").

[19] *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).

substantive legal rules.[20] 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.[21] 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, ethics are a half-measure.

D. FOCUSING ON INDIVIDUALS IS JUST PART OF THE PICTURE

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.

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.[22]

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

[20] 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.").

[21] *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.").

[22] *See* NEIL RICHARDS, WHY PRIVACY MATTERS 90-100 (2022).

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.

III. HOW TO MOVE BEYOND HALF-MEASURES

As I have explained so far, a focus merely on concepts of (1) transparency and procedure, (2) bias mitigation, (3) ethical principles, and (4) individual control is insufficient when it comes to the design and implementation of accountable AI systems. This is especially true where dangerous, disruptive systems are being released on the world by for-profit companies with scant regard to the potential larger societal effects produced by these systems.

I recommend that Congress should take three steps to go beyond AI half-measures. First**,** lawmakers must accept that AI is not neutral and take the design of AI systems seriously. Second, Congress should focus on substantive interventions such as loyalty and other duties that limit abuses of power. Finally, Congress should resist the inevitability narrative by imposing bright-line rules and outright prohibitions where appropriate.

A. ACCEPT THAT AI IS NOT NEUTRAL

As a first step in determining how AI should be regulated, I urge lawmakers to ***resist the idea that AI systems are simply neutral conduits.*** A common misconception about technologies is that they are value-neutral. 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.[23]

B. FOCUS ON SUBSTANTIVE INTERVENTIONS THAT LIMIT ABUSES OF POWER

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.[24] 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 I can offer some good news to the committee here: 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.

To that end, I recommend that since so many risks of AI systems come from within *relationships* where people are on the bad end of an information 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. 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 duty of loyalty to people who are made vulnerable to actors using AI systems would be a revolution in regulating technology as the most direct way to target business models

[23] *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).

[24] *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.

that betray people's trust.[25] These duties can support our collective well-being by incorporating civil rights principles and a focus on the social good. Building a framework around loyalty could also be a powerful rallying cry to center our new AI legislative frameworks around an anti-betrayal rule. It's hard to get inspired by waves of bureaucratic jargon about impact assessments and the specifics of neural networks, but every person on earth unfortunately knows what it feels like to be betrayed. Duties like these are proven and effective.

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.[26] I join the groups Accountable Tech, AI Now, and the Electronic Privacy Information Center in their call for bright-line rules.[27] 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. 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. A

[25] *See generally* Neil Richards & Woodrow Hartzog, *A Duty of Loyalty for Privacy Law*, 99 WASH. U. L. REV. 961; Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431 (2016); Neil Richards & Woodrow Hartzog, *Privacy's Trust Gap: A Review*, 126 YALE L.J. 1180 (2017); Woodrow Hartzog & Neil Richards, *Trusting Big Data Research*, 66 DEPAUL L. REV. 579 (2017); Neil Richards & Woodrow Hartzog, *A Relational Turn for Data Protection?,* 6 EUR. DATA PROT. L. REV. 492 (2020); Woodrow Hartzog & Neil Richards, *The Surprising Virtues of Data Loyalty*, 71 EMORY L.J. 985 (2021); Woodrow Hartzog & Neil Richards, *Legislating Data Loyalty*, 97 NOTRE DAME L. REV. REFLECTION 356 (2022).

[26] 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 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.

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

new federal agency could coordinate these protective efforts across various legal frameworks. Licensing is one 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 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.[28]

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.

C. RESIST THE INEVITABILITY NARRATIVE

This brings me to my final point, which is that I 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 (including many of the people in this room) 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.[29]

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

[28] *See generally* Gianclaudio Malgieri & Frank Pasquale, *From Transparency to Justification: Toward Ex Ante Accountability for AI*, 712 BROOK. L. SCH. LEGAL STUD. (2022).
[29] 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.").

buzz word that masks a thousand sins. My co-author Neil Richards has criticized tech company's repeated invocation of "innovation" as a canard.[30] 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.[31] 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.

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.[32] But it's dangerous to always assume the virtues of astonishingly powerful AI systems. Professor Evan Selinger and I have argued that some AI systems like face surveillance technologies are too dangerous to ever be safely deployed.[33] I recognize there is room for debate on this topic, but my point here is that when lawmakers go straight to putting up guardrails, they fail to ask the existential question about whether particular AI systems should exist at all, and under what circumstances it should ever be developed or deployed. 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.[34]

Ignoring the existential question about AI systems dooms us to a framework of half measures. We've already seen evidence of half-hearted 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

[30] *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.").

[31] *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.").

[32] 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.

[33] *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).

[34] *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.

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. CONCLUSION

The shortcomings of AI half-measures should not be read to mean that transparency, bias mitigation, ethics, and personal control are not worthwhile measures to pursue. Again, this is not to say that procedural protections are wholly ineffective; rather, it is to say that procedural protections must be backed by substantive rules that promote human flourishing.

To properly rise to the challenge of AI technologies, we need a bold, comprehensive, substantive strategy of regulation rather than resignation. Lawmakers must resist the urge to simply stack procedural half-measures on top of one another and declare the AI problem solved. Transparency, bias mitigation, ethical principles, and individual control are merely band aids for the digital lacerations inflicted by unregulated and unconstrained AI systems. Lawmakers must wrestle with the difficult questions of determining what human values and goals constitute fair, trustworthy, and accountable AI. The procedural approach has been tried in the privacy context for the last twenty-five years, and it has been a spectacular failure.[35] Now that the stakes are even higher, it is time to try a different strategy that focuses on power and substance.

In my work with Professor Neil Richards concerning trust and loyalty in the context of privacy law, I have written about the need for "a substantive embrace of a broad array of human values over privacy law's reflexive deference to individual choice, consent, and control."[36] As AI systems are increasingly integrated into our lives, whether consumer-facing or operating in the background, we must flip the presumption that burdens society with the risk of AI systems and instead place the obligation on those who develop these tools, deploy them, and profit from them to serve society.

This does not mean that lawmakers should reflexively ban AI technologies *en masse*. Rather, lawmakers should restrict the implementation and design of AI systems where it is inappropriate and use substantive regulations like duties to encourage the design and implementation of AI systems in ways which embrace human values and promote human flourishing where it is appropriate.

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

[35] *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) https://www.regulations.gov/comment/FTC-2022-0053-1071, (Comment ID: FTC-2022-0053-1071).

[36] Woodrow Hartzog & Neil Richards, *Privacy's Constitutional Moment and the Limits of Data Protection*, 61 B.C.L. REV. 1687 (2020).

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.