Florida Law Review

Volume 75 | Issue 6 Article 2

November 2023

The Law of AI for Good

Orly Lobel
University of San Diego School of Law

Follow this and additional works at: https://scholarship.law.ufl.edu/flr

Recommended Citation
Orly Lobel, *The Law of AI for Good*, 75 Fla. L. Rev. 1073 (2023).
Available at: https://scholarship.law.ufl.edu/flr/vol75/iss6/2

This Article is brought to you for free and open access by UF Law Scholarship Repository. It has been accepted for inclusion in Florida Law Review by an authorized editor of UF Law Scholarship Repository. For more information, please contact rachel@law.ufl.edu.

THE LAW OF AI FOR GOOD

Orly Lobel[*]

Abstract

Legal policy and scholarship are increasingly focused on regulating technology to safeguard against risks and harms, neglecting the ways in which the law should direct the use of new technology, particularly artificial intelligence (AI), for positive purposes. This Article pivots the debates about automation, finding that the focus on AI wrongs is descriptively inaccurate because it undermines a balanced analysis of the benefits, potentials, and risks involved in digital technology. Further, the focus on AI wrongs is normatively and prescriptively flawed, as it narrows and distorts the law reforms currently dominating tech policy debates. The Law-of-AI-Wrongs focuses on reactive and defensive solutions to potential problems while obscuring the need to proactively direct and govern increasingly automated and datafied markets and societies. By analyzing a new Federal Trade Commission report, the Biden Administration's 2022 AI Bill of Rights, President Biden's Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, and American and European legislative reform efforts—including the Algorithmic Accountability Act of 2022, the Data Privacy and Protection Act of 2022, the European General Data Protection Regulation, and the new European Union Draft AI Act—this Article finds that governments are developing regulatory strategies that almost exclusively address the risks of AI while paying short shrift to its benefits. The policy focus on the risks of digital technology is based on logical fallacies and faulty assumptions, especially when failing to evaluate AI in comparison to human decision-making and the status quo. This Article presents a shift from the prevailing absolutist approach to one of comparative cost-benefit. The role of public policy should be to oversee digital advancements, verify capabilities, and scale and build public trust in the most promising technologies.

---

* University Professor, Warren Distinguished Professor of Law, and Director of the Center for Employment and Labor Policy (CELP), University of San Diego School of Law. For thoughtful comments and conversations, I thank On Amir, Rachel Arnow-Richman, Jordan Barry, Rebecca Crootof, Thomas Kadri, Mark Lemley, Elizabeth Pollman, Nicholson Price, Christopher Slobogin, Mila Sohoni, Benjamin Van Rooj, and participants at workshops at the University of Washington, University of San Diego, Tel-Aviv University, Bar-Ilan University, University of Pennsylvania, University of Southern California, University of Colorado, Tilburg University, Yale University, AALS, the Future of Privacy Forum, and the Practicing Law Institute. For excellent research assistance, I thank Arabelle Franco, Karli Kendal, Pouch Liang, Brandee McGee, Teresa Morin, Dana Tsuri-Etzioni, and Samantha Webster. Sasha Nuñez and Elizabeth Parker provided superb legal research and library support.

Published by UF Law Scholarship Repository, 2023

A more balanced regulatory approach to AI also illuminates tensions between current AI policies. Because AI requires better, more representative data, the right to privacy can conflict with the right to fair, unbiased, and accurate algorithmic decision-making. This Article argues that the dominant policy frameworks regulating AI risks, which emphasize the right to human decision-making ("human-in-the-loop") and the right to privacy ("data minimization"), must be complemented with new corollary rights and duties: a right to automated decision-making ("human-out-of-the-loop") and a right to complete and connected datasets ("data maximization"). Moreover, a shift to proactive governance of AI reveals the necessity for behavioral research on how to establish not only trustworthy AI, but also human rationality and trust in AI. Ironically, many of the currently proposed legal protections conflict with existing behavioral insights on human-machine trust. This Article presents a blueprint for policymakers to engage in the deliberate study of how irrational aversion to automation can be mitigated through education, private-public governance, and smart policy design.

INTRODUCTION

"[W]e must admit that the Earth, the sun, the moon, the ocean and all other things are not unique, but number in numbers beyond number."[1]

– Lucretius, The Nature of Things

In the past decade, legal policy and scholarship have focused on regulating technology to safeguard against risks and harms. Policymakers and scholars have given far less attention to the ways in which the law should direct the use of digital technology, particularly artificial intelligence (AI), for positive purposes. This Article argues that the focus on AI wrongs is descriptively inaccurate because it undermines a balanced analysis of the benefits, potentials, and risks involved in automation. Further, the focus on AI wrongs is normatively and prescriptively flawed, narrowing and distorting policies currently dominating law reform debates. The "Law-of-AI-Wrongs" focuses on reactive and defensive solutions to potential problems while obscuring the need to proactively direct and govern increasingly automated and datafied markets and societies. Logical fallacies and flawed assumptions pervade the policy focus on the risks and failures of digital technology,

1. RICHARD POWER, BEWILDERMENT (2021) (quoting LUCRETIUS, DE RERUM NATURA).

which fails to consider new technologies in comparison to alternative decision-making methods and the status quo.

This Article advocates a course correction away from contemporary tech regulation's outsized and counterproductive focus on AI wrongs. Rather than devoting attention almost exclusively to preventing technology-driven risks, policymakers should refocus on how public governance can harness technology to serve social goals such as fairness, equality, welfare, health, and justice. This "Law-of-AI-for-Good" would capture the vast potential of AI while restraining its downsides. It would replace the prevailing absolutist approach that pervades contemporary policy debates with a comparative analysis of the costs and benefits of AI.

Current policy frameworks regulating AI risks present the right to human decision-making ("human-in-the-loop") and the right to privacy ("data minimization") as the primary solutions that will safeguard the public against the dangers of technology. But these approaches are becoming increasingly unrealistic and normatively flawed. If we seriously examine the mandate to consider AI's potential, the law should contemplate, under certain conditions, new corollary rights and duties. These include a right to automated decision-making ("human-out-of-the-loop") and a right to complete and connected datasets ("data maximization"). Moreover, a more balanced regulatory approach to AI reveals tensions between current AI policies. Because AI needs more representative and better data to perform accurately and fairly, and to detect inequities, too much data protection can impede the very issues that the technology needs to overcome. The right to privacy can conflict with the right to fair, unbiased, and accurate decision-making. Finally, a shift to proactive governance of AI illuminates the necessity for more behavioral research on how to establish not only trustworthy AI, but also human rationality and trust in automation. Policymakers must study and engage in policy experimentation regarding how irrational aversion to automation can be mitigated through education and design.

The need for this regulatory shift to a Law-of-AI-for-Good is particularly critical at this moment. Governments are poised to double down on regulatory strategies that nearly exclusively address the risks of AI, while paying short shrift to its benefits. In their oversight of technological advancement, the Biden Administration and the European Union (EU) are marching in lockstep to regulate the perceived harms of AI. On both sides of the pond, lawmakers are devoting their attention to

addressing the fear that algorithmic decision-making can result in errors, biases, intrusions, and exclusions.[2]

In the United States, there are currently two major AI bills before Congress. The Algorithmic Accountability Act of 2022 takes a risk-regulation approach that focuses on potential AI harms and biases.[3] The Bill, in its preamble, declares that "there are currently insufficient safeguards to protect Americans from companies' use of these programs that can exponentially amplify safety risks, unintentional errors, harmful bias and dangerous design choices."[4] It therefore prescribes investigation into the need for "any guard rail for or limitation on certain uses or applications of the automated decision system or augmented critical decision process, including whether such uses or applications ought to be prohibited or otherwise limited . . . ."[5] The American Data Privacy and Protection Act of 2022 (ADPPA) would further strengthen privacy protections through "[d]ata minimization provisions that limit data collection, use, and sharing, and that impose heightened restrictions on sensitive data such as browsing history, location data, health information, and biometric data."[6]

These legislative reforms resonate with the Biden Administration's recent policy statements and actions. In October 2023, President Biden issued a sweeping Executive Order on Safe, Secure and Trustworthy Artificial Intelligence.[7] The Executive Order focuses primarily on serious risks that the most powerful AI models might pose to national security and public safety, including the risks of engineering biological weapons

---

2. *See, e.g.*, Margot E. Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. (forthcoming 2023) (manuscript at 42, 76), http://dx.doi.org/10.2139/ssrn.4195066 [https://perma.cc/G74D-NFPV] (noting that "both the EU and the United States regulators now characterize the regulation of AI systems as risk regulation" and suggesting even stronger approaches that would take more precautionary bans and limits as opposed to risk management approaches); *see also* Gianclaudio Malgieri & Frank Pasquale, *From Transparency to Justification: Toward* Ex Ante *Accountability for AI* (Brussels Priv. Hub Working Paper, Paper No. 33, 2022), http://dx.doi.org/10.2139/ssrn.4099657 [https://perma.cc/78SR-Y7HF] (proposing "a system of 'unlawfulness by default' for AI systems, an ex-ante model where some AI developers have the burden of proof to demonstrate that their technology is not discriminatory, not manipulative, not unfair, not inaccurate, and not illegitimate in its legal bases and purposes").

3. *See generally* H.R. 6580, 117th Cong. (2022).

4. *Id.*; *Algorithmic Accountability Act of 2022*, RON WYDEN, U.S. SENATOR FOR OREGON, https://www.wyden.senate.gov/imo/media/doc/2022-02-03%20Algorithmic%20Accountability%20Act%20of%202022%20One-pager.pdf [https://perma.cc/UE8E-KC66].

5. H.R. 6580, 117th Cong. § 4(a)(6).

6. Letter from Access Now et al., to Nancy Pelosi, Speaker, U.S. House of Representatives, RE: Move H.R. 8152, the American Data Privacy and Protection Act (Aug. 25, 2022), https://cdt.org/wp-content/uploads/2022/08/Privacy-Org-Pelosi-Letter-8-25-22.pdf [https://perma.cc/WN7S-23VF]; S. 3572, 117th Cong. § 4(a)(3)(A) (2022).

7. Exec. Order No. 14110, 88 Fed. Reg. 75191 (Nov. 1, 2023) [hereinafter Biden Executive Order].

and AI enabled fraud.[8] It further calls for safeguarding Americans' privacy, preventing algorithmic bias and discrimination, and mitigate the harms of AI on the labor market.[9] A relatively small part of the sweeping order calls for the standardization of AI best practices and investment in AI research and development.[10] In October 2022, the White House released its "Blueprint for an AI Bill of Rights," which focuses on algorithmic harms and sets its core principles as data privacy and a right to opt out from AI systems, allowing for a "human fallback" instead.[11] In September 2022, the Biden Administration announced "core principles" for tech platform accountability, focused on increasing both privacy and platform liability for online harms by reforming the famous shield provided by section 230 of the Communications Decency Act of 1996.[12] These core principles, although drafted vaguely—like many other abstract calls for reform—further emphasize the risk of AI bias and call for an end to "discriminatory algorithmic decision-making."[13]

The risk-management approach to technology policy in the United States closely resembles the EU's recent reforms. The General Data Protection Regulation (GDPR), adopted in 2016, provides strong digital privacy protections and limitations on data collection and use.[14] The EU Artificial Intelligence Act (EU Draft AI Act) is a draft regulation that would ban certain uses of AI that create "unacceptable risks" and impose general limitations on the use of all AI applications.[15] All four

8. *Id.*

9. *Id.* at 75192–93.

10. *Id.* at 75196.

11. OFF. OF SCI. & TECH. POL'Y, EXEC. OFF. OF THE PRESIDENT, BLUEPRINT FOR AN AI BILL OF RIGHTS: MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE 5–7 (2022), https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf [https://perma.cc/4YBB-U7ME].

12. *Readout of White House Listening Session on Tech Platform Accountability*, WHITE HOUSE BRIEFING ROOM (Sept. 8, 2022), https://www.whitehouse.gov/briefing-room/statements-releases/2022/09/08/readout-of-white-house-listening-session-on-tech-platform-accountability/ [https://perma.cc/PL4R-NYGU] [hereinafter WHITE HOUSE BRIEFING ROOM]. During his presidential race, President Joe Biden campaigned to "revoke" section 230; in September 2022, "revoke" changed to "reform." Oliver Knox, *Biden Calls for Changing Big Tech Moderation Rules. But Not How.*, WASH. POST (Jan. 12, 2023, 11:52 AM), https://www.washingtonpost.com/politics/2023/01/12/biden-calls-changing-big-tech-moderation-rules-not-how/ [https://perma.cc/SB74-AWUG].

13. *See* WHITE HOUSE BRIEFING ROOM, *supra* note 12.

14. *See* Commission Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) art. 22, 2016 O.J. (L 119) 1, 49 [hereinafter GDPR].

15. *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, at 38–39, COM (2021) 206 final (Apr. 21, 2021) [hereinafter EU Draft AI Act].

centerpieces of AI legislation in the United States and Europe—as well as the contemporary policy statements by the executive branches—treat human decision-making as the gold standard and presumptive lawful default while limiting the reach of automation and data collection.

Of course, preventing algorithmic bias and misuse of sensitive data are important goals. However, these are only two of many goals tech policy can and should accomplish. Moreover, even the two desiderata of equality and privacy may conflict with each other in concrete policy decisions. Our current tech policy is thin and flat. It conceals that, while such normative tensions have always been a part of democratic regimes, we can steer technology's course to mitigate such conflicts between normative values. Digital technology is already gaining comparative advantage over humans in detecting discrimination; making more consistent, accurate, and nondiscriminatory decisions; and addressing the world's thorniest problems: climate, poverty, injustice, literacy, accessibility, speech, health, and safety. The role of public policy should be to oversee these advancements, verify capabilities, and build public trust of the most promising technologies. The imbalance in the contemporary tech regulation approach to AI as risk has limited these roles of public governance. Beyond safeguarding against potential risks, social democracies would benefit tremendously by setting their sights on harnessing AI for good.

This Article proceeds as follows. Part I first describes the contemporary "techlash" against AI—a mindset and regulatory framework that regards new technological capabilities as presumptively and primarily harmful. The techlash has brought mounting negative coverage of AI systems and books with titles like *Weapons of Math Destruction*,[16] *Automating Inequality*,[17] *Technically Wrong*,[18] *The New Jim Code*,[19] *Algorithms of Oppression*,[20] and *Surveillance Capitalism*.[21] Concerns about technology's failures are not unfounded, but they frequently involve distorted analyses and lead to limited, and even wrong, policy conclusions. Part I then argues that tech policy proposals often suffer from several fallacies: absolutism versus comparison; demanding

16. CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016).

17. VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2018).

18. SARA WACHTER-BOETTCHER, TECHNICALLY WRONG: SEXIST APPS, BIASED ALGORITHMS, AND OTHER THREATS OF TOXIC TECH (2017).

19. RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE (2019).

20. SAFIYA UMOJA NOBLE, ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM (2018).

21. SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER (2019).

perfection or lack of failure; engaging in the wrong comparisons; thinking of AI as static; ignoring scarcity and scale; privileging the status quo; thinking in binary solutions—adopt or ban; and making false distributional assumptions. Part I lastly demonstrates how a 2022 Congress-commissioned Federal Trade Commission (FTC) report on using AI to tackle online harms exhibits each of these fallacies about automation.

Part II introduces what AI-for-Good can look like. AI is making leaps in a wide range of areas, including environmental and climate protection, poverty alleviation, health and medicine, accessibility and accommodation, education, public governance, and law enforcement. This Part is not intended to evaluate the AI systems in each emerging technology policy area. Indeed, this Article calls on policymakers to engage in such rigorous evaluative oversight. Rather, this Part aims to offer a fuller lens that considers opportunities and advancements from which to frame our future debates and research about regulating AI. It illustrates how the contemporary law reforms discussed in the next Part, which focus on preventing the risks of AI, are limited. A fuller analysis of AI policy must include not only the risks of automation but also the counter risks and costs of *not* adopting AI to tackle pervasive social problems.

Part III describes how existing technology law reforms, as well as legal scholarship, largely focus on the risks from excess automation and data collection. These reforms thereby present the limited solutions of safeguarding against automation and protecting against surveillance and data extraction.

The pathologies of contemporary technology policy may be iterations of larger pathologies of liberal democracies and particularly the American civil rights tradition: a focus on law-as-negative-constraints rather than governance; a focus on rights as civil liberties as opposed to socioeconomic welfare; a focus on anti-classification as opposed to substantive equality and distributive justice; a focus on the individual as the unit of protection as opposed to the collective good; a focus on adaptive as opposed to anticipatory regulation; and a focus on protecting the status quo as opposed to planning for and investing in change.[22]

These broader traditions emerge in two key solutions that reign over the field of technology regulation: the twin imperatives of human decision-making and privacy—including the adoption of absolute bans

22. *See* Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 U. MIA. L. REV. 9, 9 (2003); Orly Lobel, *The Paradox of Extralegal Activism: Critical Legal Consciousness and Transformative Politics*, 120 HARV. L. REV. 937, 948 (2007); Orly Lobel, *The (Re)New(ing) Democracy and Cyclical Forms and Substance of Regulatory Governance*, YALE J. ON REG. (Aug. 2, 2022), https://www.yalejreg.com/nc/symposium-novak-new-democracy-10/ [https://perma.cc/VP7K-JDTT].

on certain technologies, primarily biometric technologies such as facial recognition. But given the potential of AI, we need to contemplate corollary rights and duties when comparative advantage is clear. These new AI-for-Good policies would include the right to automation and the right to fuller data collection.

Moreover, Part III demonstrates the internal inconsistencies in the reasons for pushing back against digitization, automation, and algorithmic decision-making. Data gaps can be particularly harmful to more vulnerable communities, disadvantaged groups, marginalized identities, and low-income individuals. Proactive, rather than reactive, regulation that mandates fuller data collection and automated processes can play a significant role in anti-discrimination policy. Part III goes on to further examine promising developments that can lead to proactive regulation, such as the EU concept of regulatory sandboxing and the CHIPS and Science Act of 2022,[23] which includes investment in infrastructure and testbeds.[24]

Finally, Part III suggests the need for studying the interactions between humans and machines. The emerging experimental literature on the trust, and distrust, of AI can serve as a blueprint for policy research and interventions. Indeed, Part III demonstrates that existing research insights should raise doubt about recent policy reforms, such as laws requiring real-time consumer notification about the use of automated processes. Just as behavioral research first developed in relation to marketing and consumer behavior and only later came to be recognized as significant in policymaking, so too should policymakers turn their attention to understanding the human biases that lead to irrational algorithmic aversion and algorithmic adoration. To support the governance of AI-for-Good, policy should aim at spurring the right amount—and the correct kind—of AI trust.

I. TECHLASH'S AUTOMATION FALLACIES

A. *What is the Techlash?*

The techlash has been described as the "growing animus toward large technology companies (a.k.a., 'Big Tech') and to a more generalized opposition to modern technology itself, particularly innovations driven by information technology."[25] While the techlash started as a backlash

23. Pub. L. No. 117-167, 136 Stat. 1366 (codified as amended in scattered sections of 15 U.S.C.).

24. *Id.* at 1570, 1585.

25. ROBERT D. ATKINSON ET AL., INFO. TECH. & INNOVATION FOUND., A POLICYMAKER'S GUIDE TO THE "TECHLASH"—WHAT IT IS AND WHY IT'S A THREAT TO GROWTH AND PROGRESS (2019), https://www2.itif.org/2019-policymakers-guide-techlash.pdf [https://perma.cc/X6WT-BQNT].

against tech companies, particularly Big Tech,[26] the mindset of skepticism, fear, and overall dystopia is now aimed at digital technologies in general, not just the companies that develop them. A 2019 report states, "[T]he techlash has created a mob mentality, and the mob is coming for innovation."[27] One can see this shift in the increasingly critical media portrayal of tech since the 2010s, as opposed to the 1980s and 1990s when the public saw tech as a force of progress and empowerment, especially for marginalized communities.[28] Most recently, the techlash centers on the dangers of AI, broadly defined as automated systems, techniques, and algorithms that perform functions—cognition, action, or emotion—traditionally performed by humans, all of which are becoming more integrated with nearly every aspect of our lives.[29]

In an August 2022 article titled *We Need to Talk About How Good A.I. Is Getting*, *New York Times* tech reporter Kevin Roose captured the

---

26. *See* Adrian Wooldridge, *The Coming Tech-lash*, ECONOMIST (Nov. 18, 2013), https://www.economist.com/news/2013/11/18/the-coming-tech-lash [https://perma.cc/CRF8-K5PL]. Oxford Dictionary defines techlash as the "strong and widespread negative reaction to the far-reaching power and influence of large technology companies." *Techlash*, OXFORD ENGLISH DICTIONARY (3d ed. 2021). But again, the current techlash is not simply a negative reaction to Big Tech—the negative is not a big-tech-lash, but more broadly fear and aversion of new technologies and their applications. As this Article shows, its policy iterations are not simply in the competition policy field, but rather they aim at regulating the risks of the technology itself. Ironically, such increased regulation may actually contribute to market concentration. *See* Orly Lobel, *The Law of the Platform*, 101 MINN. L. REV. 87, 93 (2016); Kenneth A. Bamberger & Orly Lobel, *Platform Market Power*, 32 BERKELEY TECH. L.J. 1051, 1061 (2017).

27. Atkinson et al., *supra* note 25, at 1.

28. The media reporting is particularly alarmist in recent years. *See, e.g.*, Scott Galloway, *Silicon Valley's Tax-Avoiding, Job-Killing, Soul-Sucking Machine*, ESQUIRE (Feb. 8, 2018), https://www.esquire.com/news-politics/a15895746/bust-big-tech-silicon-valley/ [https://perma.cc/329J-6RVT]. Popular documentaries like The Social Dilemma similarly sound the alarm about technology. *See* THE SOCIAL DILEMMA (Exposure Labs 2020); *see also* DOUG ALLEN, INFO. TECH. & INNOVATION FOUND. WHY SO SAD? A LOOK AT THE CHANGE IN TONE OF TECHNOLOGY REPORTING FROM 1986 TO 2013, at 2 (2017), http://www2.itif.org/2017-why-so-sad.pdf [https://perma.cc/46BN-FU96] (explaining that media coverage of technology has shifted in recent years to highlight the potential negative effects of technology rather than the positive ones).

29. *See, e.g.*, ORLY LOBEL, THE EQUALITY MACHINE: HARNESSING DIGITAL TECHNOLOGY FOR A BRIGHTER, MORE INCLUSIVE FUTURE 3–4 (2022); Dilmurod Rakhmatov & Fasliddin Arzikulov, *Prospects for the Introduction of Artificial Intelligence Technologies in Higher Education*, 11 ACADEMICIA 929, 930 (2021); Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 404 (2017). The EU Draft AI Act defines AI as

> (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods.

EU Draft AI Act, *supra* note 15, at annex I.

contemporary mindset of the techlash and conversations surrounding tech policy:

> It's a cliché in the A.I. world, to say things like "we need to have a societal conversation about A.I. risk." There are already plenty of Davos panels, TED talks, think tanks and A.I. ethics committees out there, sketching out contingency plans for a dystopian future. What's missing is a shared, value-neutral way of talking about what today's A.I. systems are actually capable of doing, and what specific risks and opportunities those capabilities present.[30]

We can retrieve this missing dialogue about AI when we adopt a Law-of-AI-for-Good approach. The following Parts will describe beneficial uses of AI today, new developments on the horizon, and how public policy can direct ever-improving AI technology for public good.

It bears repeating and emphasizing that many of the contemplated risks of automation and its disruptive power—inaccuracy, manipulation, concentration of power, job loss, exclusion, hacking, and security breakdowns—are real and significant. The issue is not whether we should be concerned with tech wrongs, tech risks, or tech fails: the answer is clearly yes. The issue is whether the concerns are either unpacked, nuanced, concrete, and balanced, or whether they are bundled, blunt, abstract, at times overstated, and effective at shaping the conversation in distorted and counter-productive ways. This Article argues the true issue is the latter. Just as understating the risks of technology is problematic, so is an exaggerated, myopic focus. Tech dystopia may be an overcorrection to tech utopia. But it suffers equally from reasoning fallacies that translate into policy blind spots.

B. *Automation Fallacies*

The first step to understanding the limits of AI-as-Wrongs policy is to unpack the flaws in common debate patterns. This Section suggests a non-exhaustive set of automation analysis flaws. Many of these fallacies stem from the same global fallacy: failure to engage in comparative analysis.

1. Fallacy 1: The Human/Machine Double Standard

The most significant overarching fallacy that the techlash lens presents is demanding AI perfection rather than comparative advantage over human decision-making and the status quo. Policy debates too often point to problems with AI as conclusory evidence of its infancy, danger,

30. Kevin Roose, *We Need to Talk About How Good A.I. is Getting*, N.Y. TIMES (Aug. 24, 2022), https://www.nytimes.com/2022/08/24/technology/ai-technology-progress.html [https://perma.cc/A87V-FP2E].

and unreadiness for public use—or AI being "rudimentary," to use the term from the FTC report analyzed below.[31] Policymakers should instead engage in comparative advantage analysis between existing systems and opportunities that automation presents. The comparisons should acknowledge the Kantian idea of "ought implies can," comparing performance to what is currently possible to achieve by humans or machines, not to a non-existent ideal.[32] For example, expecting autonomous vehicles to drive with zero crashes is less useful (and, indeed, a riskier path) than comparing human driving with self-driving vehicles to determine the relative benefit. We need to critically consider the limits and risks of both human and algorithmic decision-making. On all fronts—looking at safety, fairness, equality,[33] transparency,[34] and efficiency—we must take a comparative approach between automation and competing options, using consistent principles, metrics, and standards.

2. Fallacy 2: AI as Static & Fixed

A recurring pattern in the techlash discourse is pointing to an automated system that failed once or has been proven to be limited or biased and therefore concluding that the technology in its entirety is a failure. AI, and in particular machine learning—a system that can learn and adapt beyond explicit one-shot advance instructions by humans—is, by definition, an evolving, improving technology.[35]

Researching the literature on AI failures uncovers a handful of examples that have become iconic in the charge against AI and are retold frequently. One is the study by MIT researcher Joy Buolamwini that found facial recognition systems had the lowest accuracy when analyzing darker skinned women.[36] Another is the story of Amazon's hiring

31. *See* FED. TRADE COMM'N, COMBATTING ONLINE HARMS THROUGH INNOVATION 5 (2022), https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf [https://perma.cc/7EZE-Q9W2].

32. *See* BRITANNICA, *Ought Implies Can* (May 4, 2018), https://www.britannica.com/topic/ought-implies-can [https://perma.cc/SC24-2SYC].

33. *See* LOBEL, *supra* note 29.

34. *See, e.g.*, John Zerilli et al., *Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?*, 32 PHIL. & TECH. 661, 664 (2019).

35. *See* Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 89 (2014).

36. *See* Steve Lohr, *Facial Recognition Is Accurate, if You're a White Guy*, N.Y. TIMES (Feb. 9, 2018), https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html [https://perma.cc/N7LT-ZJSJ]; Karen Hao, *AI Is Sending People to Jail—and Getting It Wrong*, MIT TECH. REV. (Jan. 21, 2019), https://www.technologyreview.com/s/612775/algorithms-criminal-justice-ai/ [https://perma.cc/XUL4-UCC2] (discussing the increased use of predictive AI in policing and research indicating racial bias from such algorithms); Larry

algorithm that turned sexist due to training data that reflected past sexist hiring and employment patterns of the company.[37] However, the critical outcomes of the algorithm are often excluded from the story. The fact that Amazon tested for bias and never deployed this hiring system, and the fact that résumé parsing and automated screening in the job market is highly prevalent and often comparatively more predictive and inclusive than human screening, often get lost in the retelling.[38]

Ethical choices should of course be embedded in the initial design of AI systems, rather than only being introduced as down-the-line fixes or afterthoughts. Training data should be representative and inclusive. Biases in our existing social structures should inform algorithmic design and direct a departure from such past inequities. Yet, fixes are significant, and improvement matters. Technology provides opportunities to learn and correct over time. Moreover, concerns about AI inaccuracy or bias often center on AI capability as a narrow singular process—such as a single algorithm—rather than a combination of technologies, as well as machine-human interactions. In practice, the field of AI has made significant strides in combining algorithms to increase AI trustworthiness.[39] For example, scholars now consider a combination of

---

Hardesty, *Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems*, MIT NEWS (Feb. 11, 2018), https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212 [https://perma.cc/95MZ-FH8P]. For the original study, see Joy Adowaa Buolamwini, Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers (Aug. 10, 2017) (M.S. thesis, Massachusetts Institute of Technology), https://dspace.mit.edu/bitstream/handle/1721.1/114068/1026503582-MIT.pdf?sequence=1&isAllowed=y [https://perma.cc/555Q-J4CA].

37. Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 10, 2018, 7:04 PM), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [perma.cc/MVE6-9CUE].

38. *See* Danielle Li et al., *Hiring as Exploration* 25–26 (Nat'l Bureau of Econ. Rsch., Working Paper No. 27736, 2020), https://www.nber.org/system/files/working_papers/w27736/w27736.pdf [https://perma.cc/XS7C-JSJU] (comparing different types of supervised and unsupervised learning in hiring algorithms in relation to their diversity outcomes). *See generally* LOBEL, *supra* note 29 (discussing how algorithmic decision-making can overcome biases in human decision-making).

39. *See, e.g.*, Michael P. Kim et al., *Multiaccuracy: Black-Box Post-Processing for Fairness in Classification*, PROC. OF THE 2019 AAAI/ACM CONF. ON AI, ETHICS, AND SOC'Y 247, 248–52 (2019), https://dl.acm.org/doi/pdf/10.1145/3306618.3314287 [https:// perma.cc/Z3XE-XFNQ]; Maranke Wieringa, W*hat to Account for When Accounting for Algorithms: A Systematic Literature Review on Algorithmic Accountability*, PROC. 2020 CONF. ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 1, 5 (2020), https://dl.acm.org/doi/pdf/10.1145/3351095.3372833 [https://perma.cc/WRG6-ZYQZ]; Niva Elkin-Koren, *Contesting Algorithms: Restoring the Public Interest in Content Filtering by Artificial Intelligence*, 7 BIG DATA & SOC'Y 811 (2020); Cynthia Dwork et al., *Fairness Through Awareness*, PROC. 3D INNOVATIONS IN THEORETICAL COMPUT. SCI. CONF. 214, 216–23 (2012), https://dl.acm.org/doi/pdf/10.1145/209 0236.2090255 [https://perma.cc/WU5L-S6HX].

multiple algorithms an engineering gold standard in tackling AI bias.[40]

Related to the fallacy of AI as static is the fallacy of pointing to the reality that problems persist despite the availability of a technology, and then using that as evidence that technology does not work. Such reasoning disregards the question of whether an available technology has been scaled and become widespread, replacing less efficient processes. This logic is also present in the FTC report discussed below.[41]

3. Fallacy 3: Ignoring Scarcity

People often view comparisons through an apples-to-oranges lens, ignoring scarcity and accessibility of human specialists. In medicine, for example, headlines describe how an already quite great AI radiology screening algorithm performs slightly worse than—or the same as—two highly trained human radiologists performing the same screening.[42] This comparison between the accuracy of a new technology and two trained professionals working together is not the realistic comparison that most patients will face in their healthcare options.[43] Scarcity—including the challenges of global aging and elderly care, illiteracy and hunger, access to education and professional advice, and the scales of data that public and private actors need to collect, analyze, and moderate—must be part of any correct analysis of the costs and benefits of automation. The current dialogue surrounding AI underappreciates the potential of AI to help alleviate social ills where humans are short-staffed. Underappreciating that the growing demand for automated processes will require AI's assistance in governing these processes—for example, as is discussed in Section I.C, the necessity of automating content moderation—is a fallacy in itself.[44]

---

40. *See* sources cited *supra* note 39.

41. *See* FED. TRADE COMM'N, *supra* note 31.

42. *See* PRANAV RAJPURKAR ET AL., DEEP LEARNING FOR CHEST RADIOGRAPH DIAGNOSIS: A RETROSPECTIVE COMPARISON OF THE CHEXNEXT ALGORITHM TO PRACTICING RADIOLOGISTS 7–9 (PLOS MED., Nov. 2018), https://journals.plos.org/plosmedicine/article/file?id=10.1371/journal.pmed.1002686&type=printable [https://perma.cc/38YQ-FEFV].

43. *See* Maciej A. Mazurowski, *Do We Expect More from Radiology AI than from Radiologists?*, 3 RADIOLOGY: A.I., July 2021, at 2–3, https://pubs.rsna.org/doi/epdf/10.1148/ryai.2021200221 [https://perma.cc/3SQP-6G54]. Similarly, initial comparisons about the performance of adjudicative processes of humans versus machines were done in experimental settings where participants were given information about their accuracy along the way, artificially boosting the human performance results (and still resulting in similar COMPAS/human performance). Zhiyuan "Jerry" Lin et al., *The Limits of Human Predictions of Recidivism*, 6 SCI. ADVANCES no. 7, 2020, at 4–5, https://www.science.org/doi/10.1126/sciadv.aaz0652 [https://perma.cc/F86D-88J2]; *see also* W. Nicholson Price II, *Medical AI and Contextual Bias*, 33 HARV. J. L. & TECH. 65, 101–04 (2019) (arguing that human-in-the-loop solutions rely on skilled providers who often will not be present).

44. *See* Cary Coglianese & Orly Lobel, *Public Administration in the Digital Age*, *in* OXFORD HANDBOOK OF DIGITAL CONSTITUTIONALISM (forthcoming) (on file with author).

4. Fallacy 4: Risks Loom Larger than Gains

A well-established cognitive failure in the behavioral literature is loss aversion, or the perception that losses loom larger than gains. In contemporary tech policy debates, risks of overuse and underuse of an existing tool are not discussed symmetrically. Rather, risks of using available AI tools loom larger than the failure to use them. The policy perspective of AI-as-Wrongs discounts the costs of forgoing available technology and amplifies the risks of adoption. Take again the example of a radiology screening algorithm. The media is quick to raise questions about the accuracy of such systems. Yet we hardly see articles raising questions about why a professional community or a regulatory agency has not approved an automated system as the safest medical process.[45] This is also true in the policy reforms underway. For example, as will be discussed in Part III, the EU Draft AI Act contemplates the risks of using AI but is silent about the risks of not using state-of-the-art AI if it proves better than existing systems.[46]

To further underscore this point, the argument is not that the risks of adoption of AI systems are not real or even that we need to agree that they are overstated. Rather, the argument is that there are risks and costs of *not* using the most advanced technologies, just as there are risks in adopting them.

This fallacy in particular relates to the later discussion about the primacy of certain negative rights that have become dominant in tech reform. Contemplating solely the risks of automation has led to regulatory solutions that are primarily proscriptive: do not extract data, do not trace, do not adopt an imperfect algorithm, and do not automate. This protective regulatory stance exemplifies a bias in favor of inaction and the status quo that, in some instances, likely no longer serves us.

5. Fallacy 5: Thinking in Binary–Adopt or Ban–Solutions

Another recurring fallacy in tech regulatory reforms is debating the desirability of a new technology in the binary paths of permitting technology versus banning it.[47] For example, facial recognition and

45. *See, e.g.*, STAN BENJAMENS ET AL., THE STATE OF ARTIFICIAL INTELLIGENCE-BASED FDA-APPROVED MEDICAL DEVICES AND ALGORITHMS: AN ONLINE DATABASE 2 (Sept. 11, 2020), https://www.nature.com/articles/s41746-020-00324-0 [https://perma.cc/9E8U-3MCT] (calling awareness to the importance of regulatory bodies and lack of clarity around the approval of artificial intelligence and machine learning medical devices).

46. *See* EU Draft AI Act, *supra* note 15, at 1.

47. Indeed, in the scholarly literature there are voices calling for even stronger regulatory approaches than AI-as-Risk: stronger ex-ante precautionary bans and default illegality. For example, Professor Margot Kaminski explains that the contemporary AI-as-Risk approach to regulation makes "three largely unexamined policy choices around AI systems: to construct harms

biometric data collection are frequently discussed in terms of all-out bans rather than in terms of data governance and regulating misuse.[48] The Algorithmic Accountability Act of 2022 would require an assessment of the need for "guard rail[s] for or limitation[s] on certain uses or applications of the automated decision system."[49] Content moderation law reform is discussed in the narrow terms of imposing platform liability and mandating takedowns or absolute liability shields—or worse, a prohibition on private content moderation—rather than publicly governing the very broad actual private market systems already in place.[50]

Not only does the false dichotomy of "adopt or ban" overlook a sea of possibilities for directing technology for good, but it also misconstrues the reality of technology in general—once it is out there, use spreads. Indeed, a related discrepancy between the techlash debates and reality is that, as Rob Walker wrote in a 2019 *New York Times* article, while surveys about the public fears of digital systems, tech platforms, and algorithms mount, the usage and widespread adoption of new tech only continues to rise.[51] This is true even when we look at the most "backlashable"—as Walker calls them—platforms and technologies, including social networks such as Facebook, Twitter, and TikTok; voice recognition; chatbots such as GPT; and digital personal assistants.[52]

6. Fallacy 6: Distributional Assumptions

Finally, a recurring global and deeply problematic fallacy in conversations about digital technology is that this technology is not only persistently biased, but that its bias and exclusion disproportionately

---

as risks, to use risk regulation rather than precaution, and to use a particular model of risk regulation." Kaminski, *supra* note 2, at 4. Kaminski describes a precautionary approach to AI that would be even more limiting—limiting or banning certain technologies or uses as opposed to conducting risk assessments and requiring the mitigation of AI harms. *See id.*

48. *See, e.g.*, Kate Conger et al., *San Francisco Bans Facial Recognition Technology*, N.Y. TIMES (May 14, 2019), https://www.nytimes.com/.2019/05/14/us/facial-recognition-ban-san-francisco.html [https://perma.cc/9PUJ-QNQE]; Anabelle Roy, Note, *Ready Or Not Congress, Here It Comes: The Expansion Of Facial Recognition Technology Makes Its Way Into Police Practices*, 75 FLA. L. REV. 583, 583 (2023) (advocating for "a permanent ban on the use of [facial recognition technology] by law enforcement agencies.").

49. H.R. 6580, 117th Cong. § 4(6) (2022).

50. *See supra* text accompanying note 12. On September 16, 2022, the U.S. Court of Appeals for the Fifth Circuit decided *NetChoice v. Paxton*, upholding the constitutionality of a Texas law restricting large social media platforms to moderate content. *See* 49 F.4th 439, 494 (5th Cir. 2022). The U.S. Court of Appeals for the Eleventh Circuit held a similar Florida law unconstitutional. *See* NetChoice, LLC v. Att'y Gen., Fla., 34 F.4th 1196, 1231 (11th Cir. 2022).

51. Rob Walker, *There Is No Tech Backlash*, N.Y. TIMES (Sept. 14, 2019), https://www.nytimes.com/2019/09/14/opinion/tech-backlash.html [https://perma.cc/DH5B-CLEA].

52. *Id.*

harm the most vulnerable, thereby deepening inequality. In my book *The Equality Machine*, I examine this question across a range of fields of automation: hiring, promotion and pay, credit and lending, health and medical care, media and political participation, dating and intimate relations, law enforcement and public benefits.[53] My research finds that often—although of course not always—the adoption of AI supports, *or can support*, the goals of equality.[54]

In particular, the assertion that data collection is especially risky for minorities, women, and vulnerable communities is often simply wrong. The use of data collection can inevitably be positive or negative. Legal regimes can be good and evil. Alas, when women's reproductive rights are under new vicious attacks, there is a rising fear that governments or private actors may use any kind of tracking and tracing to subvert these rights.[55] Yet, even here and now, in the context of reproductive justice, the overreliance on privacy as the key to protecting women reveals the weakness of the negative liberties lens. A more positive discourse about equality, health, bodily integrity, economic rights,[56] and self-determination[57] would move us beyond the reedy looking glass of what is and is not included in privacy. As I recently described in a lecture about *Dobbs v. Jackson Women's Health Organization*, abortion rights are far more than privacy rights—they are equality rights, human rights, health rights, dignity rights, and economic rights.[58] Data under certain circumstances can serve dystopian modes of social control. But within a richer construction of what democratic liberal regimes and markets entail, and against the backdrop of long histories of offline exclusion, discrimination, uneven access, and ongoing biases and inequities, we have much to gain from automated data collection and algorithmic decision-making.

The distributional assumption fallacy and its kin automation fallacies have elevated privacy, human decision-making, and bans on risky technology above other social goals. Unchecked assumptions about harms have led to defensive and protective regulatory defaults: banning

---

53. *See generally* LOBEL, *supra* note 29.

54. *Id.*

55. *See, e.g.*, Jeremy Kahn, *After* Roe*, Fears Mount About A.I.'s Ability to Identify Those Seeking Abortions*, FORTUNE (June 28, 2022, 12:13 PM), https://fortune.com/2022/06/28/after-roe-v-wade-fear-of-a-i-surveillance-abortion/ [https://perma.cc/3ZK5-KJ3E].

56. *See, e.g.*, Andrea Flynn & Susan R. Holmberg, *America Needs Economic Rights. Now is the Time to Push for Them*, THE NATION (Jan. 11, 2019), https://www.thenation.com/article/archive/franklin-roosevelt-economic-bill-rights/ [https://perma.cc/3WUM-XAA6].

57. *See, e.g.*, Gonzales v. Carhart, 550 U.S. 124, 168 (2007).

58. Orly Lobel, *The Future of* Roe*, Reproductive Rights, and Work Equality - My Comments at Today's USD Event*, PRAWFSBLAWG (May 26, 2022, 5:20 PM), https://prawfsblawg.blogs. com/prawfsblawg/2022/05/the-future-of-roe-reproductive-rights-and-work-equality-my-comments -at-todays-usd-event.html [https://perma.cc/XLZ4-82RR].

certain technologies; limiting data collection; demanding notification, explainability, and contestation regarding the use of an automated system; and requiring reversion to human decision-making. Some of these policies provide important protections. Yet these protections are limited. Moreover, some of these reforms may in fact undermine the goal of equality. As discussed below, proactive reform proposals—such as scaling success, mandating the compilation of missing data sets, subsidizing and procuring state-of-the-art innovation, requiring the adoption of digital systems, investing in bias bug bounty systems and AI competitions, and creating AI sandboxes and testbeds—are alarmingly rare.[59]

C. *Case Study: The 2022 FTC Report on Using AI to Tackle Online Harms*

An example of a protective mindset, exemplifying some, if not all, of the AI-as-Wrongs automation fallacies described above is the 2022 FTC Report on Using AI to Tackle Online Harms.[60] The 2021 Appropriations Act "directed the [FTC] to study . . . whether and how [AI] 'may be used to identify, remove, or take any other appropriate action necessary to address' a wide variety of specified 'online harms.'"[61] Congress also asked the FTC to recommend policies and procedures for using AI to combat these online harms and any legislation to "advance the adoption and use of AI."[62] A year later, the FTC issued an eighty-two-page report.[63] The report surveys a range of private, public, and publicly funded AI tools and automated systems to tackle online harms along with offline harms that can be tackled through online systems.[64]

In effect, the report has a treasure trove of AI-for-Good examples. The meaty middle of the report that describes a range of new opportunities is sandwiched, however, by conclusory statements in the introduction and recommendations warning against—indeed strongly discouraging—the development of and reliance on automation, instead calling for safeguards against the risks of AI.[65] The report does not use the framework of comparative advantage and skirts over many issues of human decision-making in the context of tackling online harm, including inaccuracy, scarcity, cultural variation, and emotional harm to human moderators

59. *See infra* Part III.
60. *See* FED. TRADE COMM'N, *supra* note 31, at 1.
61. *Id.*
62. *Id.*
63. *Id.*
64. *See generally id.* (discussing such tools and systems).
65. *Id.* at 3, 5, 72–73, 78.

examining sensitive content.[66] Most importantly, the report does not offer a path forward for how to systematically develop, invest, or collaborate with private industry; nor does it proffer methods for monitoring, evaluating, or scaling the use of AI to tackle online harms—the original question Congress posed.

The report begins with a global warning and skepticism about technology: "Greed, hate, sickness, violence, and manipulation are not technological creations, and technology will not rid society of them."[67] After eighty-something pages, the report concludes with a similar, even more poetic apprehension: "'Platforms dream of electric shepherds,' says Tarleton Gillespie, expressing skepticism that automation can replace humans in addressing harmful online content. Legislators and regulators with similar dreams should remain skeptical as well. . . . AI is no magical shortcut."[68] These prosaic warnings and calls for skepticism would perhaps be benign if they did not foreshadow what is absent in the report.

The report's recommendations section does begin with an acknowledgment that AI is here to stay and grow.[69] The sentence that immediately follows begins with "but" and warns against misuse, over-reliance, poor results, and AI doing more harm than good.[70] The concluding takeaway that follows is that policymakers should focus on "appropriate safeguards":

> The development and deployment of automated tools to address online harms will continue with or without federal encouragement. But misuse or over-reliance on these tools can lead to poor results that can serve to cause more harm than they mitigate. For this reason, Congress, government agencies, platforms, scientists, and others should focus on appropriate safeguards.[71]

Clearly, what these appropriate safeguards *are* should be embedded in the inquiry itself as Congress directs. It stands to reason that asking about employing AI to fight online harms should include questions such as how to evaluate the strengths of the system adopted; what its comparative advantages are; what goals, outcomes, costs, benefits, and normative tradeoffs need to be decided upon; and how to differentiate between accurate and inaccurate claims about AI performance. The report cites some valuable sources that could have aided in answering these

66. *See, e.g.*, Complaint at 10–11, Scola v. Facebook, Inc., No. 18-CIV-05135 (Cal. Super. Ct. July 14, 2021) (class action by human moderators on the emotional harm inflicted upon human moderators when sifting through scores of difficult images).

67. FED. TRADE COMM'N, *supra* note 31, at 2.

68. *Id.* at 78.

69. *See id.* at 38.

70. *Id.*

71. *Id.*

questions, but the FTC's analysis does not elevate, elaborate, or extract principles for law reform. Examples from the meaty middle are glossed over quickly and dwarfed by the warnings about AI risks. The most prescriptive statement in the report is that "Congress should generally steer clear of laws that require, assume the use of, or pressure companies to deploy AI tools to detect harmful content."[72] The report concludes that:

> [S]uch tools are rudimentary and can result in bias and discrimination. Further, laws that push platforms to rapidly remove certain types of harmful content may not survive First Amendment scrutiny in any event, as they would tend both to result in the overblocking of lawful speech and impinge on platform discretion.[73]

Demonstrating the fallacies of ignoring scarcity, AI-human double standard, AI as static and fixed, and the privileging of the status quo, the FTC report points to persistent problems, such as online hate speech, as evidence that the AI tools are likely ineffective.[74] After conflating private tech efforts with scaled government human enforcement, as well as procurement, public investment in, and public commission of best practices—all of which are currently non-existent—the report concludes that Congress should not adopt any law that requires or even incentivizes companies to automate certain processes.[75] This faulty reasoning in the report continues with the argument that when there is a likelihood of a constant arms race—a game of cat and mouse or whack-a-mole—government should *steer clear* (as stated in the report's conclusion) rather than double up its efforts to match—or rather, win—said arms race.[76]

The FTC report actually points to the many layers of governance that policymakers could put in place to tackle online harms and acknowledges that tech policy often has had a tunnel vision: "With the intense focus on the role and responsibility of social media platforms, it is often lost that other private actors—as well as government agencies—could use AI to address these harms."[77] Still, the report's conclusions heavily dilute this insight.

The FTC press release in conjunction with the report becomes even more imbalanced and unabashedly averse to AI-for-Good law. The title and the subtitles read: *FTC Report Warns About Using Artificial Intelligence to Combat Online Problems: Agency Concerned with AI*

72. *Id.* at 75.

73. *Id.*

74. *Id.* at 25–26.

75. *Id.* at 5, 9.

76. *Id.* at 6, 75; *see also infra* note 281 (discussing "data poisoning" and system vulnerability).

77. FED. TRADE COMM'N, *supra* note 31, at 39.

Harms Such As Inaccuracy, Bias, Discrimination, and Commercial Surveillance Creep.[78]

Media coverage of the report has been similarly telling. For example, the Federal Reserve Bank of Atlanta described the FTC report as "a sobering report . . . [that] explores the current limits of artificial intelligence."[79] In an illuminating example of the techlash effect—an echo chamber of alarmism—this same Atlanta source continues to describe—and exaggerate—how the report lists principles for applying automated systems: "1. Human intervention is vital. When using automated tools, humans can prevent the sorts of unintended consequences that—at their most extreme—went on with the computer Hal, who very nearly murdered his human handlers in *2001: A Space Odyssey*."[80] If someone is only reading the Atlanta coverage, they would think that the Hollywood dystopian automation alarm of an AI killing its human designers appears in the original FTC report. It does not.

II. THE POTENTIAL OF AI FOR GOOD

Blame falls on automation for everything from car crashes to market crashes, from perpetuating prejudices to deepening discrimination. Again, accidents and harms are real. The purpose of this Part is not to refute AI fallibility or to evaluate any particular AI-for-Good development or application. Rather, this Part underscores the value in acknowledging the many advancements currently happening and recognizing that AI-for-Good is here.

---

78. Press Release, Fed. Trade Comm'n, FTC Report Warns About Using Artificial Intelligence to Combat Online Problems (June 16, 2022), https://www.ftc.gov/news-events/news/press-releases/2022/06/ftc-report-warns-about-using-artificial-intelligence-combat-online-problems [https://perma.cc/MJE5-SLYQ]. Also telling is that the Report was approved 4-1, with the dissenting Commissioner Noah Phillips stating that the report did not consult outside experts as Congress asked and was primarily self-referential. *Id.*; Noah Joshua Phillips, Comm'r, Fed. Trade Comm'n, Dissenting Statement Regarding the Combatting Online Harms Through Innovation Report to Congress (June 16, 2022), https://www.ftc.gov/system/files/ftc_gov/pdf/Commissioner%20Phillips%20Dissent%20to%20AI%20Report%20%28FINAL%206.16.22%20noon%29_0.pdf [https://perma.cc/X9Q3-NZFN]. Phillips described how the report "reads as a general indictment of the technology itself." *Id.* This echo chamber of fears, with the same stories and alarms about AI harms cited, retold and amplified (even after the technology has evolved), is a pattern of the techlash.

79. Claire Greene, *AI Is No Silver Bullet in Fighting Fraud*, FED. RSRV. BANK OF ATLANTA (Aug. 15, 2022), https://www.atlantafed.org/blogs/take-on-payments/2022/08/15/ai--no-silver-bullet-in-fighting-fraud [https://perma.cc/MYN4-GG8D]; *see also* Davina Garrod et al., *FTC Report on AI: A Cautionary Tale: Combatting Online Harms Through Innovation*, JD SUPRA (July 5, 2022), https://www.jdsupra.com/legalnews/ftc-report-on-ai-a-cautionary-tale-9510833/ [https://perma.cc/HB2K-NAEM] (summarizing the FTC's report to Congress concerning the limitations of AI).

80. Greene, *supra* note 79.

What is AI-for-Good? The answer depends, of course, on our definition of "good," but there are social values and goals that are likely to garner a broad consensus: protecting the environment, combatting hunger and illiteracy, advancing medicine and healthcare, and supporting education and accessibility. And of course, tackling all these social goals with increased speed, efficiency, and consistency. This Part presents several areas where there have been significant advances in addressing social challenges through the adoption of automated processes.

Public policy should be ready to research and evaluate AI, support experimentation, and scale the best of it. Importantly, we do not have to agree on whether a particular technology is currently safer or fairer than non-automated systems to agree that we need to have a conversation about what happens tomorrow (or in five years) when an even better version of that technology exists.

A. *Environmental/Climate Applications*

There are numerous applications for AI when it comes to environmental efficiency and climate change mitigation. Organizations now use AI in climate modeling, predicting weather and wind power, adjusting turbines and propellers to maximize efficiency, and decentralizing energy grids to optimize energy storage and use.[81] An oft-cited problem with solar and wind energy solutions is intermittency: energy can only be collected when the sun or wind contacts the mechanism.[82] AI can learn to constantly move propellers to the ideal position according to wind and weather patterns to optimize energy storage and usage.[83] Organizations also use AI to predict storms, heat waves, power outages, fires, lightning strikes, and grid failures before they happen—turning utility systems into proactive rather than merely reactive mechanisms.[84] NASA reportedly used AI to track Hurricane Harvey with far more accuracy than former models.[85]

In the quest for clean oceans, the nonprofit organization The Ocean Cleanup, in collaboration with Microsoft's AI for Earth initiative, has developed a machine learning system that tracks plastic pollution and directs technologies to remove plastic from the oceans.[86] Around nine

81. *See* Amy L. Stein, *Artificial Intelligence and Climate Change*, 37 YALE J. ON REG. 890, 900–07 (2020).

82. *Id.* at 901–02.

83. *Id.* at 902.

84. *Id.* at 910–13.

85. Chris Milliner et al., *Tracking the Weight of Hurricane Harvey's Stormwater Using GPS Data*, 4 SCI. ADVANCES, no. 9, 2018, at 4–5, https://www.science.org/doi/10.1126/sciadv.aau2477 [https://perma.cc/RAL6-MZX4].

86. MICROSOFT, *AI for Earth Partners*, https://www.microsoft.com/en-us/ai/ai-for-earth-the-ocean-cleanup [https://perma.cc/HHX8-6TTX].

million tons of trash and debris end up in the ocean every year.[87] At this pace, plastic could overtake the fish population by 2050.[88] Dying fish, depleted seafood supply, and rising sea levels due to global warming have all disrupted the ecological balance dramatically.[89] Another nonprofit, OceanMind, uses satellite data and AI to identify patterns of overfishing by tracking the movement of boats and ships and comparing the data to past activities.[90] Wildlife applications offer immense promise as well, greatly improving, for example, anti-poaching surveillance and prevention of power-mill-induced bird mortality.[91] Another promising development is employing AI to tackle the problem of deforestation. Using satellite images, the image processing algorithm can identify unofficial roads, thereby predicting future deforestation.[92]

AI can help prepare and support communities dealing with natural disasters. Mor Schlesinger, head of engineering at Google Crisis, describes how Google spends millions of dollars on her team.[93] Every time a crisis mode is adopted, Google halts the display of advertisements for particular searches and instead displays only essential information.[94] The idea for Google Crisis came from Google engineers themselves.[95] In 2010, fires spread in Israel's Carmel Mountains.[96] Yossi Matias, head of

87. *Id.*

88. *Will There Be More Plastic Than Fish in the Sea?*, WORLD WILDLIFE FUND-UK, https://www.wwf.org.uk/myfootprint/challenges/will-there-be-more-plastic-fish-sea [https://perma.cc/6M2P-VM7P].

89. *See Effects of Climate Change on Ecology*, UNIV. CORP. FOR ATMOSPHERIC RSCH., https://scied.ucar.edu/learning-zone/climate-change-impacts/ecology [https://perma.cc/9G5E-B3Z3].

90. *See* Alex Thornton, *How AI and Satellites are Used to Combat Illegal Fishing*, MICROSOFT (June 6, 2019), https://news.microsoft.com/on-the-issues/2019/06/06/ocean-mind-illegal-fishing/ [https://perma.cc/UY7T-28AY].

91. *See, e.g.*, Enrico Di Minin & Christoph Fink, *How Machine Learning Can Help Fight Illegal Wildlife Trade on Social Media*, THE CONVERSATION (Apr. 23, 2019, 9:59 AM), https://theconversation.com/how-machine-learning-can-help-fightillegal-wildlife-trade-on-social-media-115021 [https://perma.cc/AV4A-XEZ5]; Julio Hernandez-Castro & David L. Roberts, *Automatic Detection of Potentially Illegal Online Sales of Elephant Ivory via Data Mining*, PEERJ COMPUT. SCI. (2015), https://doi.org/10.7717/peerj-cs.10 [https://perma.cc/9JQG-WCB6]; Molly Espey & Eamon Espey, *Using Markets to Limit Eagle Mortality from Wind Power*, PERC (July 26, 2022), https://www.perc.org/2022/07/26/using-markets-to-limit-eagle-mortality-from-wind-power/ [https://perma.cc/35SR-8XJK].

92. *Imazon – Microsoft AI for Earth*, MICROSOFT, https://www.microsoft.com/en-us/ai/ai-for-earth-imazon [https://perma.cc/9GLZ-MR4Y].

93. *Sounding the Alarm – Using AI in Disaster Management*, GOOGLE, https://about.google/intl/ALL_us/stories/soundingthealarm/ [https://perma.cc/YJR4-5HBN].

94. *See* Omer Kabir, *From Crisis Response to Accessibility Tools: Some Recent Developments From Google's Israel R&D Center*, CTECH (May 19, 2019, 5:55 PM), https://www.calcalistech.com/ctech/articles/0,7340,L-3762427,00.html [https://perma.cc/A92T-UBVD].

95. *Id.*

96. *Id.*

Google Israel's R&D Center, saw the fires raging from an office window but could not find information about them online.[97] Now, when natural disasters like fires, tornadoes, earthquakes, and hurricanes hit, Google Crisis activates public alerts to provide information to anxious Google searchers.[98] Such information includes how long the emergency is expected to last, recommended safety actions, and where to find additional resources.[99] Google Crisis uses machine learning to forecast floods and other natural disasters resulting in earlier and better advance notice to evacuate disaster-prone areas, significantly extending the window of warning and preparation.[100]

Importantly, while AI can support the creation of greener tech, AI itself is an energy-consuming technology.[101] Investing in energy efficient AI systems is key to ensuring AI has a positive net effect on sustainability. Questions about such positive net effects are not easy or straightforward, but they are the ones we need to be asking, and, inevitably, answering globally.

B. *Food Scarcity & Poverty Alleviation*

Another frontier where AI has the potential to do good is poverty alleviation. AI can help governments and nonprofits by deciphering satellite imaging to understand and forecast where resource scarcities lie. Stanford scholars are using such imaging to estimate poverty levels across communities in Africa using a model that analyzes infrastructure such as roads, agriculture, building materials, structures, and water.[102] During natural disasters, AI helps map impoverished areas to better respond to and alleviate imminent scarcity.[103] AI systems also address poverty and inequality through predictions of at-risk neighborhoods showing signs of community decline or gentrification.[104] The United

97. *Id.*

98. *Id.*

99. *Id.*

100. *Id.*

101. *See generally* KATE CRAWFORD, THE ATLAS OF AI: POWER, POLITICS, AND THE PLANETARY COSTS OF ARTIFICIAL INTELLIGENCE (2021) (explaining that the creation of AI systems depends on exploiting energy).

102. May Wong, *Stanford Researchers Harness Satellite Imagery and AI to Help Fight Poverty in Africa*, STAN. NEWS (May 22, 2022), https://news.stanford.edu/2020/05/22/using-satellites-ai-help-fight-poverty-africa/ [https://perma.cc/TZ5V-Y2U7].

103. David Mhlanga, *Artificial Intelligence in the Industry 4.0, and Its Impact on Poverty, Innovation, Infrastructure Development, and the Sustainable Development Goals: Lessons from Emerging Economies?*, 13 SUSTAINABILITY, no. 11, 2021, at 10, https://doi.org/10.3390/su131115788 [https://perma.cc/Z68W-DEKN].

104. *See, e.g.*, Shadi Copty, *The Urban Institute and IBM team up to fight inequality using AI*, IBM (Mar. 11, 2021), https://www.ibm.com/blogs/journey-to-ai/2021/03/urban-institute-and-

Nations (UN) Global Pulse, a big data and AI initiative, uses information from mobile phone purchases and anonymized call records to track poverty and direct food and health policy.[105] Researchers also use AI to analyze satellite imagery to estimate poverty levels in villages, helping governments and Non-Governmental Organizations (NGOs) create priorities in service delivery.[106] During the COVID-19 pandemic, Carnegie Mellon used AI to analyze school bus routes to determine the most optimal locations and cost-effective routes for food distribution to children who relied on school meals to fight hunger.[107] More broadly, AI can also help to map financial risk, predict financial downturns, and spot financial crimes such as money laundering.[108] In credit and mortgage decisions, AI can improve accuracy resulting in higher lender approval rates, especially for underserved applicants.[109] AI can analyze risk factors for over-indebtedness and assess poverty risk in debt-ridden nations.[110]

In the rapidly developing field of "AgTech" (agricultural technology) software developers have created AI that can inexpensively test water purity, forecast crop yields, and detect diseased crops.[111] Farmview, for

---

ibm-team-up-to-change-arming-policy-makers-with-timely-insights/ [https://perma.cc/G3HJ-7XRH]; Seth Dobrin, *Urban Institute and IBM Help Cities Measure Gentrification*, IBM BLOG (Oct. 21, 2021), https://www.ibm.com/blogs/journey-to-ai/2021/10/urban-institute-and-ibm-help-cities-measure-gentrification/ [https://perma.cc/K9KN-AXQQ]; Gabriel Gilling et al., *Predicting Neighborhood Change Using Publicly Available Data and Machine Learning* (July 30, 2021), https://ssrn.com/abstract=3911354 [https://perma.cc/YP93-UBQC].

105. *Using Mobile Phone Data and Airtime Credit Purchases to Estimate Food Security*, UN GLOBAL PULSE, https://www.unglobalpulse.org/project/using-mobile-phone-data-and-airtime-credit-purchases-to-estimate-food-security/ [https://perma.cc/4MSS-T3J4].

106. *See* Joseph Bennington-Castro, *AI Is a Game-Changer in the Fight Against Hunger and Poverty. Here's Why*, NBC (June 21, 2017, 3:26 PM), https://www.nbcnews.com/mach/tech/ai-game-changer-fight-against-hunger-poverty-here-s-why-ncna774696 [https://perma.cc/DM38-JXCM].

107. Jessica Kent, *Machine Learning Helps Reduce Food Insecurity During Covid-19*, HEALTH IT ANALYTICS (Nov. 6, 2020), https://healthitanalytics.com/news/machine-learning-helps-reduce-food-insecurity-during-covid-19 [https://perma.cc/96TP-U5FH].

108. *Can Using Software to Map Financial Risks Predict the Next Downturn?*, KNOWLEDGE AT WHARTON (Sept. 20, 2018), https://knowledge.wharton.upenn.edu/article/can-using-software-map-financial-risks-help-predict-next-downturn/ [https://perma.cc/87HA-DWKH].

109. Susan Wharton Gates et al., *Automated Underwriting in Mortgage Lending: Good News for the Underserved?*, 13 HOUS. POL'Y DEBATE 369, 370 (2002).

110. Mário Boto Ferreira et al., *Using Artificial Intelligence to Overcome Over-Indebtedness and Fight Poverty*, 131 J. BUS. RSCH. 411, 412 (2021); @pramodAIML, *How AI Can Help Alleviate Poverty?*, MEDIUM (Aug. 20, 2020), https://medium.com/predict/how-ai-ml-can-help-alleviate-poverty-917b92a72844 [https://perma.cc/RAY6-Q4R3].

111. *See* Margaret A. Goralski & Tay Keong Tan, *Artificial Intelligence and Poverty Alleviation: Emerging Innovations and Their Implications for Management Education and Sustainable Development*, 20 INT'L J. MGMT. EDUC. 1, 3 (2022) (presenting examples of AI being used to test water and identify diseased plants); Asaf Tzachor, *Using AI in Agriculture Could Boost Global Food Security–but We Need to Anticipate the Risks*, THE CONVERSATION (Mar. 29,

example, uses AI with robots to help increase the yield of certain staple crops.[112] Robots drive through the fields using cameras and lasers to measure characteristics of plants such as size, color, and possible signs of disease.[113]

C. *Health & Medicine*

In medicine, AI is already bringing dazzling positive innovation. AI can bring earlier and more accurate diagnoses; advanced imaging; better treatment and patient adherence; safer medical procedures; increased access and reduced costs of quality care; more complete, connected, and accurate datasets; and discovery of new connections between data and disease to discover novel treatments and cures.[114] In 2017, DeepMind captivated the world with its AlphaGo computer program, which beat the world's champions in Chinese Go.[115] Over a decade ago, IBM's Watson beat Kasparov in chess.[116] But powerful machines have been doing much more than beating humans in games. Watson's descendant is Watson for Oncology and AlphaGo's descendant is AlphaFold, both of which contribute to immense medical and pharmaceutical breakthroughs.[117] In medical research, 2022 saw immense leaps when DeepMind announced that AlphaFold predicted nearly all 200 million known proteins.[118] The

---

2022, 10:57 AM), https://theconversation.com/using-ai-in-agriculture-could-boost-global-food-security-but-we-need-to-anticipate-the-risks-178104 [https://perma.cc/FA8J-VRH4] (discussing the risks and benefits of using AI in agriculture); Sanjiv Sharma & Jashandeep Singh, *A Review on Usage and Expected Benefits of Artificial Intelligence in Agriculture Sector*, 29 INT'L J. ADVANCED SCI. & TECH. 1078, 1080 (2020) (identifying ways AI has been used to help farmers predict disease and forecast crop yields) .

112. *FarmView: CMU Researchers Working to Increase Crop Yield With Fewer Resources*, CARNEGIE MELLON UNIV., https://www.cmu.edu/work-that-matters/farmview [https://perma.cc/W5KB-2TTC].

113. *See id.*; Lisa Rabasca Roepe, *How AI Can Help Fight Poverty*, DELL (Nov. 14, 2018), https://www.dell.com/en-us/perspectives/how-ai-can-help-fight-poverty/ [https://perma.cc/A5LB-TY9E].

114. *See* LOBEL, *supra* note 29, at ch. 5 (presenting advancements in digital technology that can improve our health, with particular attention to underrepresented groups).

115. *Google AI Defeats Human Go Champion*, BBC (May 25, 2017), https://www.bbc.com/news/technology-40042581 [https://perma.cc/V3SM-Q9X8].

116. Dustin Waters, *The Historic Chess Showdown Between Man and AI, Decades Before ChatGPT*, WASH. POST (May 22, 2023, 7:30 AM), https://www.washingtonpost.com/history/2023/05/22/garry-kasparov-chess-deep-blue-ibm/ [https://perma.cc/H4NF-669C].

117. *See, e.g.*, Factspan Analytics Inc., *Watson (AI) For Oncology-A Thousand Case Studies*, MEDIUM (May 14, 2021), https://factspan.medium.com/watson-ai-for-oncology-a-thousand-case-studies-3e8a931d0663 [https://perma.cc/7SP2-E2U5] (discussing the use of Watson for Oncology in cancer treatment); *Alphafold Reveals the Structure of the Protein Universe*, GOOGLE DEEPMIND (July 28, 2022), https://www.deepmind.com/blog/alphafold-reveals-the-structure-of-the-protein-universe [https://perma.cc/A5SP-XTUT] (discussing the applications for AlphaFold's protein database).

118. *Alphafold Reveals the Structure of the Protein Universe*, *supra* note 117.

database will serve to develop new drugs, vaccines, and treatments.[119] *Science* named AI-powered protein prediction as its 2021 Breakthrough of the Year.[120]

AI has applications in oncology, neurology, ophthalmology, and cardiology among other areas for diagnostics, prevention, surgery, and recovery.[121] Advances in AI in radiology have already resulted in better image processing and reduced radiation doses, leading to faster, safer, and more cost-effective care.[122] AI can assist in bringing care to rural communities, overcoming language barriers, and providing follow-up services in a timelier manner through telemedicine.[123] Machine learning software that designs automated personalized messaging is also improving patient adherence to treatment and medication.[124] Medical

119. *Id.*

120. Walter Beckwith, *Science's 2021 Breakthrough: AI-Powered Protein Prediction*, AM. ASS'N FOR THE ADVANCEMENT OF SCI. (Dec. 17, 2021), https://www.aaas.org/news/sciences-2021-breakthrough-ai-powered-protein-prediction [https://perma.cc/9SQ3-CNWE]. *Science* is published by the American Association for Advancement of Science (AAAS) as a "family of journals." *See About Science & AAAS*, SCIENCE, https://www.science.org/content/page/about-science-aaas [https://perma.cc/UBV4-XFUV].

121. *See generally* Mélanie Bourassa Forcier et al., *Integrating Artificial Intelligence into Health Care Through Data Access: Can the GDPR Act as a Beacon for Policymakers?*, 6 J. L. & BIOSCIENCES 317 (2019) (discussing the applications of AI in healthcare); Price II, *supra* note 43 (presenting ways that AI can increase access to medical care); Philipp Tschandl et al., *Comparison of the Accuracy of Human Readers Versus Machine-Learning Algorithms for Pigmented Skin Lesion Classification: An Open, Web-Based, International, Diagnostic Study*, 20 LANCET ONCOLOGY 938 (2019) (presenting an empirical study demonstrating that AI can assist in the detection of cancerous lesions).

122. *See* Zvonimir Krajcer, *Artificial Intelligence in Cardiovascular Medicine: Historical Overview, Current Status, and Future Directions*, 49 TEX. HEART INST. J. 1, 6 (2022) ("An AI-guided system for acquiring cardiac magnetic resonance (CMR) images (so-called 'one-click MRI') (HeartVista, Inc.) has reduced scanning time from 90 minutes to 15 minutes. An AI-based algorithm for calculating fractional flow reserve from coronary computed tomographic angiograms (HeartFlow FFRCT) can diagnose the severity of coronary artery disease with very high sensitivity and specificity.").

123. *See generally* Hassane Alami et al., *Artificial Intelligence in Health Care: Laying the Foundation for Responsible, Sustainable, and Inclusive Innovation in Low- and Middle-Income Countries*, 16 GLOBALIZATION & HEALTH 1 (2020) (discussing scenarios where AI could promote access to healthcare); Sonu Bhaskar et al., *Designing Futuristic Telemedicine Using Artificial Intelligence and Robotics in the COVID-19 Era*, 8 FRONTIERS PUB. HEALTH 1 (2020) (presenting potential benefits of AI assisted telemedicine).

124. Keren B. Aharon et al., *Improving Cardiac Rehabilitation Patient Adherence Via Personalized Interventions*, 17 PLOS ONE, no. 8, 2022, at 1, 6; Sarah Kamensky, Note, *Artificial Intelligence and Technology in Health Care: Overview and Possible Legal Implications*, 21 DEPAUL J. HEALTH CARE L., Spring 2020, at 1, 4, 6 n.37; Shelby Engelbrecht, *Artificial Intelligence in Health Law*, MICH. TECH. L. REV. (Feb. 1, 2022). AI applications for emotional support and mental health are also a booming industry, with apps like Woebot designed to help with postpartum depression. WOEBOT HEALTH, https://woebothealth.com/ [https://perma.cc/DHF6-WHJ5]; Mallory Hackett, *Digital Chatbot Woebot Lands FDA Breakthrough Designation*

providers implement AI in emergency rooms to identify and fast-track patients needing X-rays or other tests before a medical exam.[125] By analyzing triage data and automatically ordering tests, AI cuts down wait times.[126]

Similarly, AI unquestionably played a leading role in the race for vaccines during the COVID-19 pandemic, comparing data from clinical trials, sifting through health databases, anticipating patient health risks, predicting hospital capacity, supporting the development of remote care, and tracking and tracing the spread and pace of infection.[127]

D. *Accessibility & Accommodation*

The disability community has seen great progress in public accessibility aided by technology. Speech-to-text and text-to-speech technologies, as well as facial recognition and personal digital assistants, can assist and ensure fuller, real-time participation. Technologies such as facial recognition have suffered from patterned inaccuracies, under-sampling the speech patterns of minorities and women.[128] At the same time, they have been making tremendous leaps in both accuracy and application. GnoSys, which functions as a Google translator for those with hearing and speech impairments, uses computer vision and neural networks to translate sign language and gestures into text and speech.[129] Google's DeepMind uses AI to create lip reading algorithms to interpret whole phrases.[130] Microsoft's Seeing AI is a computer vision program designed to narrate the environment to the visually impaired, again using facial recognition, as well as expression recognition technology to help

to Tackle Postpartum Depression, MOBI HEALTH NEWS (May 26, 2021, 1:36 PM), https://www.mobihealthnews.com/news/digital-chatbot-woebot-lands-fda-breakthrough-designation-tackle-postpartum-depression [https://perma.cc/2BH4-H5DT].

125. CIFAR, AI & HEALTHCARE: A FUSION OF LAW & SCIENCE 15 (2021), https://cifar.ca/wp-content/uploads/2021/03/210218-ai-and-health-care-law-and-science-v8-AODA.pdf [https://perma.cc/Y5B4-CDPU].

126. *Id.*

127. Vasiliy Andreevich Laptev et al., *Medical Applications of Artificial Intelligence (Legal Aspects and Future Prospects)*, LAWS, Dec. 29, 2022, at 2, 4, https://doi.org/10.3390/laws11010003 [https://perma.cc/5CJ8-T23K]; Barry Solaiman, *Addressing Access with Artificial Intelligence: Overcoming the Limitations of Deep Learning to Broaden Remote Care Today*, 51 U. MEM. L. REV. 1103, 1114–15 (2021).

128. LOBEL, *supra* note 29, at ch. 6.

129. *How AI Can Improve the Lives of People with Disabilities*, SMARTCLICK, https://smartclick.ai/articles/how-ai-can-improve-the-lives-of-people-with-disabilities/ [https://perma.cc/96F8-CHCG].

130. Robin Christopherson, *Lip-Reading with Google's DeepMind AI: What It Means for Disabled People, Live Subtitling and Espionage!*, ABILITYNET (Mar. 2, 2017), https://abilitynet.org.uk/lipreading-google-deepmind-future-disabled [https://perma.cc/C52K-ZR79].

describe what people look like.[131] Further, Microsoft's AI for Accessibility program awards grants to emerging AI technology that empowers people with disabilities.[132]

In the job market, more specifically, digital platforms and applications like LinkedIn are helping to make recruiting information more accessible by, for example, enabling text-to-speech to discover jobs.[133] Notably, given these developments, the doctrine of reasonable accommodation under the Americans with Disabilities Act must evolve. The costs and ability to accommodate differences dramatically change with emerging technology,[134] and regulators and adjudicators must recognize these shifts. An evolving accommodation doctrine is an example of how public policy can take more affirmative stances and elevate the standards of inclusion, expanding beyond the limited anti-discrimination disparate treatment lens.

At the intersection of health, care, and accessibility are care robots. Care robots—most recognizably in the form of a baby seal named Paro—have been approved by the Federal Drug Administration (FDA) to help alleviate social isolation and loneliness among older adults, as well as help patients with depression and other mental and physical health

131. *Seeing AI in New Languages*, MICROSOFT, https://www.microsoft.com/en-us/ai/seeing-ai [https://perma.cc/U8UT-ZUC3]. OrCam is another app that uses voice commands to help describe visual information for the blind and visually impaired. OrCam Staff, *Hey Or Cam: New Voice Assistant for People with Visual Impairment*, ORCAM (Aug. 5, 2021), https://www.orcam.com/en-us/blog/hey-orcam-new-voice-assistant-for-people-with-visual-impairment [https://perma.cc/THY6-5WD6].

132. *AI for Accessibility Grants*, MICROSOFT, https://www.microsoft.com/en-us/ai/ai-for-accessibility-grants [https://perma.cc/WB23-GCLY].

133. Ioana Tanase, *How AI is Being Used to Improve Disability Employment*, MICROSOFT ACCESSIBILITY BLOG (Jan. 13, 2022), https://blogs.microsoft.com/accessibility/how-ai-is-being-used-to-improve-disability-employment/ [https://perma.cc/C8K7-2DEE]. LinkedIn has also made its platform more accessible, accounting for light sensitivities and visual impairment. *See* Melissa Selcher, *Our Journey to Make LinkedIn More Inclusive and Accessible*, LINKEDIN (Apr. 28, 2021), https://www.linkedin.com/pulse/our-journey-make-linkedin-more-inclusive-accessible-melissa-selcher/ [https://perma.cc/FX95-C3SV]. LinkedIn also offers learning courses for employers such as Supporting Workers with Disabilities and Hiring and Supporting Neurodiversity in the Workplace. *Hiring and Supporting Neurodiversity in the Workplace*, LINKEDIN LEARNING, https://www.linkedin.com/learning/hiring-and-supporting-neurodiversity-in-the-workplace [https://perma.cc/N3EM-AZWD]; *Supporting Workers with Disabilities*, LINKEDIN LEARNING, https://www.linkedin.com/learning/supporting-workers-with-disabilities [https://perma.cc/C7ZD-7BZ4].

134. *See* Marianne DelPo Kulow & Scott Thomas, *Assistive Technology and the Americans with Disabilities Act: Endearing Employers to These Reasonable Accommodations*, 40 BERKELEY J. EMP. & LAB. L. 257, 264 (2019); *Apple Introduces New Features for Cognitive Accessibility, Along with Live Speech, Personal Voice, and Point and Speak in Magnifier*, APPLE (May 16, 2023), https://www.apple.com/newsroom/2023/05/apple-previews-live-speech-personal-voice-and-more-new-accessibility-features/ [https://perma.cc/MYS8-4KNN].

issues.[135] During the COVID-19 pandemic, the state of New York provided more than 800 older residents with a new roommate named ElliQ, a robot that initiates conversation and interaction to provide stimulation and learns from the conversations to improve its future chats.[136] The growing demand for caregivers has accelerated the development of AI-care technologies.[137] Care robots provide a constant presence for users and can be trained specifically to help patients with Alzheimer's or dementia.[138] Their use during COVID-19 pandemic proved that, in such an event, care robots can be especially valuable when human caregivers fear infection.[139]

E. *Education*

AI has growing applications in the field of education.[140] Resources, especially the most valuable of them all—teachers themselves—are frequently overloaded.[141] Educational disparities are rampant, and helping teachers in underfunded educational communities is especially valuable.[142] New technologies have the potential to reach and educate underserved children and adults around the world.[143] According to the

135. Takanori Shibata, *Therapeutic Seal Robot as Biofeedback Medical Device: Qualitative and Quantitative Evaluations of Robot Therapy in Dementia Care*, 100 PROC. IEEE 2527, 2530, 2532 (2012).

136. Arun Kristian Das, *How Robots Help Older New Yorkers Fight Social Isolation*, FOX 5 N.Y. (May 25, 2022), https://www.fox5ny.com/news/elliq-robot-companion-for-senior-citizens [https://perma.cc/M9L8-TA9Y].

137. Donna S. Harkness, *Bridging the Uncompensated Caregiver Gap: Does Technology Provide an Ethically and Legally Viable Answer?*, 22 ELDER L.J. 399, 417 (2015).

138. Adriana Krasniansky, *Exploring Elder Care Robotics: Emotional Companion Robots*, HARV. L. PETRIE-FLOM CTR. BILL OF HEALTH BLOG (Nov. 25, 2019), https://blog.petrieflom.law.harvard.edu/2019/11/25/exploring-elder-care-robotics-emotional-companion-robots/ [https://perma.cc/838E-XVK7].

139. *See* Nancy S. Jecker, *You've Got a Friend in Me: Sociable Robots for Older Adults in an Age of Global Pandemics*, 23 ETHICS & INFO. TECH. (SUPP. 1) 35, 36–37, 41 (2021).

140. *See, e.g.*, Rakhmatov & Arzikulov, *supra* note 29; YJ Yang, *A.I. Can Help Solve America's Education Crisis*, FORTUNE (July 14, 2020, 5:00 PM), https://fortune.com/2020/07/14/education-crisis-artificial-intelligence/ [https://perma.cc/99ZT-YRJP]; Chris Chambers Goodman, *Just-AIED: An Essay on Just Applications of Artificial Intelligence in Education*, 123 W. VA. L. REV. 937, 939–40, 946 (2021); JOYCE J. LU & LAURIE A. HARRIS, CONG. RSCH. SERV., ARTIFICIAL INTELLIGENCE (AI) AND EDUCATION (2018), https://crsreports.congress.gov/product/pdf/IF/IF10937 [https://perma.cc/HNP2-TD29].

141. *See, e.g.*, Jordan Bowen, *Florida Teacher Shortage Hitting Record High as Students Adjust to New School Year*, FOX 13 NEWS (Sept. 7, 2023, 10:36 PM), https://www.fox13news.com/news/florida-teacher-shortage-hitting-record-high-as-students-adjust-to-new-school-year [https://perma.cc/AW4K-8VX8].

142. Shalni Gulati, *Technology-Enhanced Learning in Developing Nations: A Review*, 9 INT'L REV. RSCH. IN OPEN & DISTRIBUTED LEARNING, no. 1, 2008, at 2, 3, https://doi.org/10.19173/irrodl.v9i1.477 [https://perma.cc/NW2C-JKGZ].

143. *See, e.g.*, *id.* at 1, 4, 6, 9, 11; Rachel Goldman et al., *Using Educational Robotics to Engage Inner-City Students with Technology*, *in* Int'l Conf. of the Learning SCI. 214, 214 (2004).

latest UN figures, only 63% of the world's population in 2021 used the Internet, notwithstanding an increase of almost 17% since 2019, with nearly 800 million people becoming online users in two years.[144] As the UN report calls it, connectivity is a "'grand canyon' separating the digitally empowered from the digitally excluded, with 96[%] of the 2.9 billion still offline living in the developing world."[145] While the digital gender divide is narrowing, "women remain digitally marginalized in many of the world's poorest countries, where online access could potentially have its most powerful effect."[146] With the help of AI satellite imagery mining, the UN International Children's Emergency Fund and other organizations are collaborating in mapping schools to target investment to areas that most need it to increase online access.[147]

At school, AI helpers can assume the task of individual one-on-one interactions and feedback, supporting the work of human educators.[148] As AI technologies further develop and improve, schools and educational organizations will likely rely on e-learning and AI to complement and even substitute exclusively in-person, human-operated education. A study reviewing the developments in AI in education from 2000 to 2019 shows the growing positive findings of AI tech on learning performance and outcomes.[149]

F. *Agency Compliance & Law Enforcement*

In earlier work on occupational safety and health regulation, I researched how scarcity of resources, agency inspectors, and budgetary constraints have significantly limited the Occupational Safety and Health Administration's (OSHA) ability to detect and deter safety risks.[150] This scarcity is true for basically every regulatory agency. Government

144. INT'L TELECOMM. UNION, MEASURING DIGITAL DEVELOPMENT: FACTS AND FIGURES 2021, https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2021.pdf [https://perma.cc/8PFW-T4NW].

145. *Id.*

146. *Id.*

147. Iyke Maduako et al., *Automated School Location Mapping at Scale from Satellite Imagery Based on Deep Learning*, 14 REMOTE SENSING, no. 4, art. 897, 2022, at 19, https://www.mdpi.com/2072-4292/14/4/897 [https://perma.cc/LK6E-G8AC].

148. LOBEL, *supra* note 29, at ch. 10 (describing the pioneering research of MIT computer scientist Cynthia Breazeal).

149. Xieling Chen et al., *Two Decades of Artificial Intelligence in Education: Contributors, Collaborations, Research Topics, Challenges, and Future Directions*, 25 EDUC. TECH. & SOC'Y 28, 28 (2022).

150. Orly Lobel, *Governing Occupational Safety in the United States*, *in* LAW AND NEW GOVERNANCE IN THE EU AND THE US 269, 271–75 (Grainne De Burca & Joanne Scott eds., 2006); Orly Lobel, *Interlocking Regulatory and Industrial Relations: The Governance of Worker Safety*, 57 ADMIN. L. REV. 1071, 1074, 1080–81 (2005) [hereinafter Lobel, *Interlocking Regulatory and Industrial Relations*].

agencies are nearly always strapped for time, staff, and money.[151] In a recent study, researchers used machine learning methods to estimate the effects of counterfactual targeting rules OSHA could deploy: that is, using AI to help the agency decide how and which workplaces to inspect.[152] The researchers estimated that "OSHA could have averted over twice as many injuries if its inspections had targeted the establishments" identified by AI, and that such AI-based inspection "regime[] would have generated over $1 billion in social value over the decade . . . examine[d]."[153]

Agencies are already increasingly using automated tools to make decisions about enforcement, caseload management, benefits, and the application of rules. Agencies as varied as the Internal Revenue Service, Bureau of Labor Statistics, Social Security Administration, Environmental Protection Agency (EPA), Securities and Exchange Commission (SEC), Federal Communications Commission (FCC), the United States Patent and Trademark Office, and the FDA employ automated systems for governmental tasks once done by humans.[154] For example, the Department of Veterans Affairs uses AI to administer veteran benefits,[155] while the Department of Education uses an automated chatbot to help navigate student loan applications.[156] The Department of Health and Human Services sponsored the creation of an AI-based tool

151. *See, e.g.*, Nick Buffie et al., *These 6 Priorities Show the Need for a Robust Domestic Discretionary Budget*, CTR. AM. PROGRESS (Feb. 11, 2022), https://www.americanprogress.org/article/these-6-priorities-show-the-need-for-a-robust-domestic-discretionary-budget/ [https://perma.cc/UW94-6GEP].

152. Matthew S. Johnson et al., *Improving Regulatory Effectiveness through Better Targeting: Evidence from OSHA* 1 (Inst. for Rsch. on Lab. & Emp., Working Paper No. 107-19, 2019), https://irle.berkeley.edu/files/2019/09/Improving-Regulatory-Effectiveness-through-Better-Targeting.pdf [https://perma.cc/JRP3-HS2S].

153. *Id.*

154. David Debarr & Maury Harwood, *Relational Mining for Compliance Risk*, IRS RSCH. BULL., 2004, at 175, 177, https://www.irs.gov/pub/irs-soi/04debarr.pdf [https://perma.cc/W3UM-266C] (discussing use of computers to perform an initial screening of tax returns); Lea Helmers et al., *Automating the Search for a Patent's Prior Art with a Full Text Similarity Search*, PLOS ONE, Mar. 4, 2019, at 1, 2 (using artificial intelligence to compare patent applications to existing applications); David A. Bray, *An Update on the Volume of Open Internet Comments Submitted to the FCC*, FED. COMMC'N COMM'N (Sept. 17, 2014, 1:02 PM), https://www.fcc.gov/news-events/blog/2014/09/17/update-volume-open-internet-comments-submitted-fcc (discussing the FCC's effort to use an automated system to sort through public comments); M. Hino et al., *Machine Learning for Environmental Monitoring*, 1 NATURE SUSTAINABILITY 583, 583 (2018) (discussing how machine-learning methods can foster "efficient use of . . . limited resources" to enforce environmental regulation).

155. *See* Press Release, U.S. DEP'T VETERANS AFFS., VA Adopts New Artificial Intelligence Strategy to Ensure Trustworthy Use of Technology for Veteran Care (Oct. 14, 2021, 10:42 AM), https://www.va.gov/opa/pressrel/pressrelease.cfm?id=5729 [https://perma.cc/LK3C-3PZC].

156. *See Meet Aidan*, FED. STUDENT AID, https://studentaid.gov/h/aidan [https://perma.cc/79TH-U9PL].

to detect illegal opioid sellers.[157] The FDA similarly began using AI in criminal investigations and to mine through online reports on unsafe food.[158]

Close to home, at the University of California, San Diego, government-funded projects are leading the way in developing AI to detect online sales of illegal drugs and illegal COVID-19 health products.[159] The range of these new tools is far-reaching; they are designed to identify and track drug sales, find the culprits along the supply chain, help identify the victims, and support victim rehabilitation.[160] At the University of Southern California, researchers have worked with Greek border officials to use machine learning to screen travelers for COVID-19, finding that the AI system detected at least twice as many—and, under certain conditions, four times as many—asymptomatic travelers than the traditional border screening.[161] In an area related to much of my recent research—the widespread practice of unenforceable clauses in consumer and employer contracts[162]—Poland's

157. *Using Artificial Intelligence Technologies to Expose Darknet Opioid Traffickers*, NAT'L INST. OF JUST. (Aug. 31, 2018), https://nij.ojp.gov/funding/awards/2018-75-cx-0032 [https://perma.cc/5XW5-4R6T]; *see also* Rebecca Heilweil, *AI Can Help Find Illegal Opioid Sellers Online. And Wildlife Traffickers. And Counterfeit*s, VOX (Jan. 21, 2020, 8:10 AM), https://www.vox.com/recode/2020/1/21/21060680/opioids-artificial-intelligence-illegalonline-pharmacies [https://perma.cc/X5KK-QAXJ] (discussing how an AI-based tool can track digital drug dealers and illegal internet pharmacies).

158. *See, e.g.*, Adyasha Maharana et al., *Detecting Reports of Unsafe Foods in Consumer Product Reviews*, 2 JAMIA OPEN 330, 331 (2019), https://doi.org/10.1093/jamiaopen/ooz030 [https://perma.cc/2V8G-VHBC].

159. Tim Ken Mackey et al., *Big Data, Natural Language Processing, and Deep Learning to Detect and Characterize Illicit COVID-19 Product Sales: Infoveillance Study on Twitter and Instagram*, 6 JMIR PUB. HEALTH & SURVEILLANCE 360, 361 (2020), https://publichealth.jmir.org/2020/3/e20794/PDF [https://perma.cc/K9PP-DNPY]; Tim Ken Mackey et al., *Twitter-Based Detection of Illegal Online Sale of Prescription Opioid*, 107 AM. J. PUB. HEALTH 1910, 1910–1911 (2017), https://ajph.aphapublications.org/doi/10.2105/AJPH. 2017.303994 [https://perma.cc/PTM7-M95Y].

160. *See* Neal Shah et al., *An Unsupervised Machine Learning Approach for the Detection and Characterization of Illicit Drug-Dealing Comments and Interactions on Instagram*, 43 SUBSTANCE ABUSE 273 (2022); Tatyana Sushina & Andrew Sobenin, *Artificial Intelligence in the Criminal Justice System: Leading Trends and Possibilities*, 441 ADVANCES IN SOC. SCI., EDUC. & HUMANS. RSCH. 432, 434 (2020), https://www.atlantis-press.com/proceedings/icseal-6-19/125940991 [https://perma.cc/X2WE-URY7]; Shubpreet Kaur et al., *Artificial Intelligence Framework for Identifying the Population Addicted to Drugs: Markov Decision Process*, 2ND INT'L CONF. ON COMPUTATIONAL METHODS IN SCI. & TECH., 2021, at 243, 243, https://ieeexplore.ieee.org/abstract/document/9784721?casa_token=F-vb2-nIpKMAAAAA:hTs [https://perma.cc/RHP9-FS2P]; Peter N. Salib, Abolition by Algorithim (unpublished manuscript) (on file with author).

161. Hamsa Bastani et al., *Efficient and Targeted COVID-19 Border Testing Via Reinforcement Learning*, 599 NATURE 108, 108 (2021).

162. *See* Orly Lobel, *Boilerplate Collusion: Clause Aggregation, Antitrust Law & Contract Governance*, 106 MINN. L. REV. 877, 882 (2021), https://papers.ssrn.com/sol3/papers.cfm

Office of Competition and Consumer Protection is reportedly developing an automated system to detect unlawful contract clauses in consumer (and potentially also employment) contracts.[163] Also, an AI worth noting in the field of competition policy is a machine learning tool developed to monitor whether Amazon displays its own brands ahead of better-known brands with higher stars and ratings.[164]

Law enforcement's use of automated systems has been an especially fraught debate. As seen in Part I, the contemporary AI discourse highlights the potential risks of such applications. As we will see in Part III, the policy reforms currently underway are designed to prevent such risks. Yet, the upsides of AI are immense. Automated decision-making is often fairer, more efficient, less expensive, and more consistent than human decision-making. Algorithms reveal patterns often unseen by humans. For example, as described above, since the inception of the COVID-19 pandemic, governments have been tracking and tracing the spread of the disease as well as fighting back against deceptive exploits of the pandemic, such as false information and the selling of products under fraudulent claims about cures and protections.[165] The advantages of computers in computing tasks are obvious: they have unparalleled speed and processing power to mine through vast amounts of data, and they are unlikely to suffer from cognitive depletion and cognitive irrationalities. As Nobel laureate Daniel Kahneman and coauthors Olivier Sibony and Professor Cass Sunstein wrote in their recent book, *Noise*, AI "can be far less imperfect than noisy and often-biased human judgment."[166] Moreover, because many industries are now using algorithms, the work of government enforcement and monitoring needs to match their speed and competency.[167] AI can help government agencies and alleviate bureaucratic burdens with tasks ranging from

?abstract_id=3810250 [https://perma.cc/G8QH-YLUT]; MARK LEMLEY & ORLY LOBEL, SUPPORTING TALENT MOBILITY AND ENHANCING HUMAN CAPITAL: BANNING NONCOMPETE AGREEMENTS TO CREATE COMPETITIVE JOB MARKETS 1, 2 (2021), https://fas.org/wp-content/uploads/2021/01/Microsoft-Word-Supporting-Talent-Mobi...mpetitive-Job-Markets_LobelLemley.pdf [https://perma.cc/A95H-EHXT]; RACHEL ARNOW-RICHMAN ET AL., SUPPORTING MARKET ACCOUNTABILITY, WORKPLACE EQUITY, AND FAIR COMPETITION BY REINING IN NON-DISCLOSURE AGREEMENTS 1, 2 (2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4022812 [https://perma.cc/BWR5-H8NG].

163. *See, e.g.*, Rajmund Molski, *Competition Law and Artificial Intelligence – Challenges and Opportunities*, 14 TEKA KOMISJI PRAWNICZEJ PAN ODDZIAŁ W LUBLINIE 339, 348, https://ojs.academicon.pl/tkppan/article/view/4533/4605 [https://perma.cc/3XAG-CC2T].

164. Julia Angwin, *The Mathematics of Amazon's Advantage*, MARKUP (Oct. 16, 2021, 8:00 AM), https://themarkup.org/newsletter/hello-world/the-mathematics-of-amazons-advantage [https://perma.cc/T4YT-R5XV].

165. Bastani et al., *supra* note 161, at 108.

166. DANIEL KAHNEMAN ET AL., NOISE: A FLAW IN HUMAN JUDGMENT 272 (2021).

167. Cary Coglianese, *Optimizing Regulation for an Optimizing Economy*, 4 U. PA. J. L. & PUB. AFFS. 1, 1–2 (2018).

assessing crime risk, spotting electoral fraud, and furthering aviation safety.[168] Automation and digitization can also alleviate the burdens of administrative paperwork, which a 2021 Presidential Executive Order describes as a burden that exceeds nine billion hours annually with regard to federal agencies.[169] The downsides of adopting the wrong AI, automating badly, and arbitrarily adopting AI systems without government guidance on best practices are also clearly immense.

III. AI-FOR-GOOD RIGHTS

A. *A Right to Automated Decision-Making*

AI is rapidly entering every domain: finance, health, work, military, agriculture, education, transportation, dispute resolution, entertainment, art, dating, and more. As algorithms become central in both private industry and government operations, the fear of their flaws has brought pushback and a loud call for maintaining human decision-making, human control, human action, human accountability, a human face, and human oversight.[170] As discussed below, not only is this right to a human-in-the-loop often inconsistent, vague, and multi-meaning, the rapid advancement of AI requires a corollary right to demand automation when such a shift results in better, safer, fairer, and more accurate outcomes.

1. Human-out-of-the-Loop versus Human-in-the-Loop

The idea of a "right to a human" can mean many different things, relate to many different stages, and imply many different rules. In a myopic way, however, a primary solution that emerges is the right to a human as either the final decision-maker or as a replacement for automated processes altogether.[171] Numerous governmental bodies have

168. Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1153–66 (2017).

169. Exec. Order No. 14058, 86 Fed. Reg. 71357, 71357 (Dec. 13, 2021).

170. *See, e.g.*, Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245, 1248 (2016); Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 651 (2020); Ryan Calo & Danielle Keats Citron, *The Automated Administrative State: A Crisis of Legitimacy*, 70 EMORY L.J. 797, 800 (2021); AI NOW INST., LITIGATING ALGORITHMS: CHALLENGING GOVERNMENT USE OF ALGORITHMIC DECISION SYSTEMS 3 (2018); Anupam Chander, *The Racist Algorithm?*, 115 MICH. L. REV. 1023, 1025, 1040 (2017); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 671, 732 (2016); Aziz Z. Huq, *Artificial Intelligence and the Rule of Law* 1 (Univ. Chi. L. Sch., Working Paper No. 764, 2021), https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?article=2194&context=public_law_and_legal_theory [https://perma.cc/3DTQ-QUW4].

171. Kiel Brennan-Marquez et al., *Strange Loops: Apparent Versus Actual Human Involvement in Automated Decision-Making*, 34 BERKELEY TECH. L.J. 745, 746–47 (2019); Meg Leta Jones, *The Right to a Human in the Loop: Political Constructions of Computer Automation*

adopted or proposed laws that accord the right to a human decision-maker. We encountered the human-in-the-loop prominently highlighted in the FTC report on tackling online harms.[172] The report's first conclusion is that human intervention is vital.[173] The EU has made the right to a human-in-the-loop a centerpiece in its technology regulation.[174] Indeed, the first ethics principle of trustworthy AI by the EU is human agency and oversight.[175] The EU Draft AI Act, which defines AI broadly to include machine learning and statistical methods more generally, includes both a right to a human decision-maker and a right to disclosures when humans are interacting with AI.[176] In Article 22 of the GDPR, which went into effect in 2018, the data subject has the default right not to face a decision made solely based on automated decision-making.[177] In California, new provisions of the California Privacy Rights Act (CPRA), which went into effect in 2023, provide users the right to opt out of automated decision-making.[178] One recent article surveying American law identifies forty-one laws and proposed reforms that require humans in the decision-making loop.[179]

Under some of these frameworks, algorithms are allowed to help humans in the decision-making process, but the subjects of the regulatory process have a right to a final decision made by a human.[180] For example, the Wisconsin Supreme Court discussed that agencies and courts must "exercise discretion when assessing a COMPAS risk score[, an automated system on sentencing/release,] with respect to each individual defendant."[181] That is, the court allowed the use of the AI system on the condition that the final decision-maker is a human who exercised

and Personhood, 47 SOC. STUDS. SCI. 216, 231–32 (2017); Adrian Bridgwater, *Machine Learning Needs a Human-in-the-Loop*, FORBES (Mar. 7, 2016, 1:00 PM), https://www.forbes.com/sites/adrianbridgwater/2016/03/07/machine-learning-needs-a-human-in-the-loop/?sh=3c233114cabf [https://perma.cc/Q8PT-U5DK].

172. FED. TRADE COMM'N, *supra* note 31.

173. *Id.*

174. *See* EU Draft AI Act, *supra* note 15.

175. *Ethics Guidelines for Trustworthy AI*, EUR. COMM'N (Apr. 8, 2019), https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html [https://perma.cc/8FRN-3T2C].

176. *See* EU Draft AI Act, *supra* note 15.

177. GDPR, *supra* note 14.

178. Shannon Yavorsky, *New U.S. State Privacy Laws Zero in on Artificial Intelligence*, ORRICK (Aug. 11, 2022), https://www.orrick.com/en/Insights/2022/08/New-State-Privacy-Laws-Zero-in-on-AI [https://perma.cc/YT7A-FHZQ].

179. Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, 45 COMPUT. L. SEC. REV. 1, 2 (2022).

180. Ben Wagner, *Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems*, 11 POL'Y & INTERNET 104, 108–09 (2019); Aziz Z. Huq, *Constitutional Rights in the Machine-Learning State*, 105 CORNELL L. REV. 1875, 1906 (2020); Jones, *supra* note 171, at 223–24.

181. State v. Loomis, 881 N.W.2d 749, 764–65 (Wis. 2016).

independent judgment in deciding whether to follow the AI's recommendation.[182]

Scholars have begun to unpack and inquire about the effectiveness of human-in-the-loop regulations. A recent article by Professors Rebecca Crootof, Margot Kaminski, and William Nicholson Price raises questions about humans in the loop as a regulatory safeguard.[183] The article warns that "[r]ather than marrying the best of [both], hybrid . . . systems can exacerbate the worst of each, while adding new sources of error" as information is lost in translation.[184] When designing autonomous vehicles (AVs), for example, systems that hand off control to the driver need to deal with the reality that drivers in autonomous vehicles lose focus and may drift off; systems might need to include "alerts and sufficient time" for the operator to take control.[185]

Yet, despite this recent questioning of the wisdom and effects of human-in-the-loop, I have yet to see stronger implications emerging from these insights. Namely, that, under certain circumstances, there should be a right to an artificial decision-maker, alongside a corollary duty to automate.[186] Put differently, there should be a prohibition on humans entering the loop when such entrance would diminish the benefits of automation and bring error and bias.

Indeed, take the example of AVs; there will come a time, and we can disagree about when that time will come, that AVs will be a lot safer than human drivers.[187] Nobel laureate Daniel Kahneman recently predicted in an interview that "[b]eing a lot safer than people is not going to be enough. The factor by which they [AVs] have to be more safe than humans is really very high."[188] That is an alarming prediction. Ever the optimist, I venture to disagree with Kahneman. Humans all over the world have already agreed that automation is a lot safer in high-stakes travel: air travel. The entirety of the international aviation industry operates with the gold standard of autopilot when weather conditions are

182. *Id.* at 769.

183. Rebecca Crootof et al., *Humans in the Loop*, 76 VAND. L. REV. 429, 508 (2023).

184. *Id.* at 438.

185. *Id.* at 439.

186. *See generally* Wesley Newcomb Hohfeld, *Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 26 YALE L.J. 710, 717 (1917) (discussing correlative rights).

187. *See* Alex John London, *Groundhog Day for Medical Artificial Intelligence*, 48 HASTINGS CTR. REP. (2018), https://doi.org/10.1002/hast.842 [https://perma.cc/RU4N-AFBV].

188. Tim Adams, *Daniel Kahneman: 'Clearly AI Is Going to Win. How People Are Going To Adjust Is a Fascinating Problem'*, The GUARDIAN (May 16, 2021), https://www.theguardian.com/books/2021/may/16/daniel-kahneman-clearly-ai-is-going-to-win-how-people-are-going-to-adjust-is-a-fascinating-problem-thinking-fast-and-slow [https://perma.cc/9GY5-9AZB].

harsh.[189] If consumers are comfortable with this standard, there is no reason to believe that we cannot learn to love "a lot safer" autonomous cars—as well as fully autonomous commercial planes. An acceptance of an "a lot safer" standard instead of a "really very high factor" of better performance will take policy, education, facts, design, and learning. It would take a cultural shift and a willingness from humans to give up control. This may involve different nudges for different generations: those who grew up driving may be more reluctant to relinquish the steering wheel, literally and figuratively. Section III.C below investigates what behavioral law of AI aversion could look like. But on the policy side, what is important to emphasize is that a right and duty to automate are not only possible but morally correct. To underscore this point, failing to acknowledge the possibility of legally prohibiting human decision-making under certain circumstances is a normative choice. It is a regulatory (in)action that may come at a serious cost.

For decades, psychologists and behavioral scientists have documented the fact that statistical models, i.e., algorithms, can outperform people, including experts such as physicians.[190] Yet, as philosopher Alex London sharply put it:

> Professionals routinely overestimate their ability to perform such tasks and underestimate the value of actuarial methods for making health care decisions. Precisely because medical diagnostic and prediction decisions are intimately bound up with matters of life and death, perpetuating the neglect of highly accurate algorithmic decision tools is not a benign deference to professional prerogative. It is a potentially lethal hubris whose tithe is exacted in avoidable morbidity and mortality.[191]

The hubris is even worse when we take a public policy perspective. Requiring humans to be the final decision-makers in high stakes processes is not only a flawed solution in contexts where AI has clearly reached comparative advantages, but it also risks perpetuating irrational fears about AI instead of helping debias citizens about the comparative risks of technology. Most troubling, such hubris ironically risks legitimizing the use of flawed algorithms rather than working to make the algorithms better because it continues the legacy of automation fallacies. A right to automation and a duty to automate would require more robust

189. *See, e.g.*, John *Cox*, *How Often Do Airline Pilots Rely on Autopilot? What Happens If a Plane's Engine Falls Off?*, USA TODAY (Jan. 25, 2022, 5:33 PM), https://www.usatoday.com/story/travel/columnist/cox/2022/01/16/how-often-do-airline-pilots-rely-autopilot/9189477002/# [https://perma.cc/E8PQ-69JT].

190. *See, e.g.*, Robyn M. Dawes et al., *Clinical Versus Actuarial Judgment*, 243 SCI. 1668, 1673 (1989).

191. *See* London, *supra* note 187.

methods to compare between, and subsequently monitor, humans, machines, and other, better, safer, newer machines. Take, for example, a requirement by an agency like the Federal Aviation Administration or the FDA to adopt automation when an AI system proves safer for travelers or patients or to adopt AI through a "standard of care" common law lens.[192] Indeed, tort law has embedded within it a concept of "state-of-the-art" safety.[193] Increasingly, such a safety standard will mean automation. It will require standardization and openness in a way that does not exist today and a far more active role of public oversight.[194]

The focus on the risks and failures of AI has obscured the need for such AI-for-Good regulation. This lens is critical moving forward. Shifting to a more balanced research and policy agenda in the rapid development of AI has profound implications in virtually every policy field and sphere of life. In particular, as considered in the next Section, it requires more complete data collection and publicly available datasets.

2. Data is Desirable to Detect Discrimination

As we saw above, a particular fear of AI that has received much attention is that of AI bias. Scholars and policymakers alike warn that because data is historically biased toward certain groups or classes, discriminatory results may still emerge from automated algorithms that are designed in racial- or gender-neutral ways. Discriminatory results can also occur even when decision-makers are not motivated to discriminate.[195] When discussing AI bias, comparative advantages are rarely the framework of the analysis. The focus is instead on extending

192. As an example of standard of care today, states already require doctors and pharmacists to consult state databases and screening tools to identify over-users and over-prescribers. *See Prescription Drug Monitoring Programs: A Guide for Healthcare Providers*, 10 SUBSTANCE ABUSE & MENTAL HEALTH SERV. ADMIN. 1, 5 (2017), https://store.samhsa.gov/sites/default/files/d7/priv/sma16-4997.pdf [https://perma.cc/FSD5-GNY6]. *But see* Jennifer D. Oliva, *Dosing Discrimination: Regulating PDMP Risk Scores*, 110 CALIF. L. REV. 47, 47 (2022) (describing how prescription drug monitoring programs such as NarxCare could artificially inflate marginalized patients' risks for prescription drug misuse).

193. *See, e.g.*, Bryan H. Choi, *Crashworthy Code*, 94 WASH. L. REV. 39, 47 (2019). Environmental policy as well the EPA, notwithstanding the current pushback from the U.S. Supreme Court, seek to require the best available technology to reduce carbon emissions. *See infra* Section III.A.3 (discussing the major questions doctrine).

194. *See also infra* Section III.C.2 (explaining different ideas researchers have proposed for reporting the use of AI through datasheets).

195. *Civil Rights Principles for Hiring Assessment Technologies*, LEADERSHIP CONF. ON CIV. & HUM. RTS. (July 29, 2020), http://civilrightsdocs.info/pdf/policy/letters/2020/Hiring_Principles_FINAL_7.29.20.pdf [https://perma.cc/83N9-546B] ("Hiring assessment technologies are one of many barriers that may impede equity and inclusion in the workforce. Artificial intelligence, by its very nature, risks replicating and deepening existing inequities when it relies on data from the current workforce that is not sufficiently representative because of historic discrimination.").

the mandate of nondiscrimination to automated processes.[196] Such extension is a necessary and crucial expansion. Yet, rarely do we contemplate a mandate to shift to an automated system if it proves to be more inclusive. Arguably, our laws already demand such positive action. As the U.S. Supreme Court, interpreting Title VII, described, Congress demanded from employers "the removal of artificial, arbitrary, and unnecessary barriers to employment when the barriers operate invidiously to discriminate on the basis of racial or other impermissible classification."[197] If we now have far stronger tools to detect and remove such arbitrary barriers, we need to use them. Already, digital tools are helping employers expand the hiring pool and diversify the workforce.[198] When this possibility of adopting better technology to support equality is raised, scholars ask about the legality of such efforts, framing them as "affirmative action."[199] But that may well be the wrong term for the wrong question.

As we shall see in Section III.B, to fight against discrimination and to ensure that algorithms are more inclusive, more data is usually needed. The need for better data persists even in the face of the AI field's trajectory toward more top-down reasoning, which signals less reliance on past data and more independent common sense.[200] The power of AI comes from its ability to use large amounts of data. AI can be brittle, biased, and fallible, or it can be increasingly reliable, unbiased, and consistent. When biases stem from partial, unrepresentative, and tainted data, the solution may be the collection of more, rather than less, data.[201] But tech policy is largely informed by privacy and anti-surveillance scholarship and activism, calling for more limits on data collection. This

196. *See id.*

197. Dothard v. Rawlinson, 433 U.S. 321, 328 (1977) (quoting Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971)).

198. *See, e.g.*, Orly Lobel, *Knowledge Pays: Reversing Information Flows and the Future of Pay Equity*, 120 COLUM. L. REV. 547, 591 (2020) [hereinafter Lobel, *Knowledge Pays*].

199. *See* Peter Salib, *Big Data Affirmative Action*, 117 NW. U. L. REV. 821, 821 (2022); Jason R. Bent, *Is Algorithmic Affirmative Action Legal?*, 108 GEO. L.J. 803, 807–08 (2020); Daniel Ho & Alice Xiang, *Affirmative Algorithms: The Legal Grounds for Fairness as Awareness*, 2020 U. CHI. L. REV. ONLINE 134, 134 (2020); Chander, *supra* note 170, at 1025.

200. *See* H. James Wilson et al., *The Future of AI Will Be About Less Data, Not More*, HARV. BUS. REV. (Jan. 14, 2019), https://hbr.org/2019/01/the-future-of-ai-will-be-about-less-data-not-more [https://perma.cc/384D-C5NS].

201. *See* Alexander Amini et. al., *Uncovering and Mitigating Algorithmic Bias Through Learned Latent Structure*, PROC. 2019 CONF. ON A.I., ETHICS, & SOC'Y 289, 291 (Jan. 27–28, 2019); W. Nicholson Price II, *Risk and Resilience in Health Data Infrastructure*, 16 COLO. TECH. L.J. 65, 78–80 (2017) (arguing for the collection of representative health data as infrastructure for innovation). *See generally* Ana Bracic et al., *Exclusion Cycles: Reinforcing Disparities in Medicine*, 377 SCI. 1158 (2022) (arguing that biased health datasets will lead to cycles of exclusion and poor performance for minority patients); LOBEL, *supra* note 29 (discussing how technology is a powerful tool that, if used well, can harness equality and ensure a better future).

is a point of tension that we must recognize as we discuss the next default regulatory reform: enhancing privacy protections.

3. Machines are Major

In 2022, the Supreme Court in *West Virginia v. EPA*,[202] in a 6-3 decision centered around carbon emissions and climate change, dealt a major blow to the EPA regulatory power.[203] The Court, expanding "the major questions doctrine," held that any time an agency decides on a "major question," the regulation is presumptively invalid unless Congress specifically authorized the regulation of the question.[204] The EPA sought to regulate coal-fired power plants, the single largest source of carbon emissions contributing to climate change.[205] The EPA set carbon limits and directed states to rely on alternative sources of energy.[206] The Court held that agencies cannot adopt rules that are transformational to the economy unless Congress specifically authorized such a rule.[207] Chief Justice John Roberts wrote that "in certain extraordinary cases, both separation of powers principles and a practical understanding of legislative intent make us 'reluctant to read into ambiguous statutory text' the delegation claimed to be lurking there."[208] In her dissent, Justice Elena Kagan explained that the Clean Air Act clearly anticipates that the regulatory agency will have to regulate new problems with new science.[209] The Court, on the other hand, according to Justice Kagan, "does not have a clue about how to address climate change," yet, it "appoints itself—instead of Congress or the expert agency—the decision-maker on climate policy. I cannot think of many things more frightening."[210] Frightening indeed is that any regulatory agency that

202. 142 S. Ct. 2587 (2022).

203. *Id.* at 2616.

204. *Id.* at 2614.

205. *Id.* at 2599; *Sources of Greenhouse Gas Emissions*, EPA, https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions [https://perma.cc/HR92-BCNS].

206. *West Virginia*, 142 S. Ct. at 2599.

207. *Id.* at 2608–09.

208. *Id.* at 2609 (citation omitted).

209. *Id.* at 2628 (Kagan, J., dissenting).

210. *Id.* at 2644; *see also* Mila Sohoni, *The Major Questions Quartet*, 136 HARV. L. REV. 262, 283 (discussing Kagan's dissenting opinion in West Virginia v. EPA); Daniel T. Deacon & Leah M. Litman, *The New Major Questions Doctrine*, 109 VA. L. REV. 1009 (2023) (analyzing the Court's application of the major questions doctrine in *West Virginia v. EPA*). *See generally* Cass R. Sunstein, *There Are Two "Major Questions" Doctrines*, 73 ADMIN. L. REV. 475 (2021) (purporting that there are two versions of the major questions doctrine with very different meanings). After *West Virginia v. EPA* was decided, Professor Sunstein resolved his previous article in favor of there likely only being one doctrine backed by "an incompletely theorized agreement in favor of the major questions doctrine, [with] two justifications [that] might lead in different directions." Cass R. Sunstein, *Two Justifications for the Major Questions Doctrine*, 76 FLA. L. REV. (forthcoming 2024).

seeks to adopt rules about measuring, disclosing, and, most importantly, mandating a critical shift to safer technology now stands on shaky ground. Imagine a regulation that directs states to automate all their transportation controls or requires that AI systems be involved in energy storage processes. Under current jurisprudence, courts could construe these as major shifts that may entail major startup costs. The vagueness of the major questions doctrine is even more alarming in an age of rapid leaps in science and technology that can aid in addressing major threats to our planet and well-being. Regulatory agencies such as the EPA, OSHA, FTC, and others should be empowered to proactively promulgate rules requiring automation, not just rules safeguarding against it.

B. *A Right to Data Collection*

1. Against Privacy's Privilege

In the contemporary climate of AI-as-Wrongs, privacy stands as the focal point of activists and policymakers who aim to push back against datafication—the private and public collection and processing of data. Professor Shoshana Zuboff has been influential in coining the term "surveillance capitalism," describing the "unilateral claiming" of "human experience as free raw material for translation into behavioral data."[211] The FTC, HIPAA, California's Consumer Privacy Act, and other central regulations each aim to protect against the extraction and sharing of consumer data. A new bill before Congress, the American Data Privacy and Protection Act, seeks to further strengthen privacy protection and impose restrictions on data collection.[212] The Act includes broad definitions, covering any entity that collects, processes, or transfers covered data and is subject to the jurisdiction of the FTC, including nonprofits, telecommunications, and common carriers.[213] It covers all "information that identifies or is linked or reasonably linkable . . . to an individual or device . . . linkable to an individual."[214] The "data minimization" section of the draft bill imposes "a baseline duty on all covered entities not to unnecessarily collect or use covered data in the

211. Shoshana Zuboff, *You are Now Remotely Controlled*, N.Y. TIMES (Jan. 24, 2020), https://www.nytimes.com/2020/01/24/opinion/sunday/surveillance-capitalism.html [https://perma.cc/H3Y6-5MDF]; Shoshana Zuboff, *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization*, 30 J. INFO. TECH. 75, 79 (2015); Shoshana Zuboff, *Surveillance Capitalism and the Challenge of Collective Action*, 28 NEW LAB. F. 10, 14 (2019) (discussing the evolution of surveillance capitalism). See generally ZUBOFF, *supra* note 21, for a more extensive analysis of surveillance capitalism and the quest by powerful corporations to predict and control human behavior.

212. H.R. 8152, 117th Cong. pmbl. (2022).

213. *Id.* at § 2(9).

214. *Id.* at §§ 8(A), 29(A).

first instance, regardless of any consent or transparency requirements."[215] The bill prohibits "collecting, processing, or transferring covered data beyond what is reasonably necessary, proportionate, and limited to provide specific products and services requested by individuals . . . or for a purpose expressly permitted by the Act."[216] The Biden Administration, in its newly released *Core Principles for Platforms*, endorses the bill and calls for:

> clear limits on the ability to collect, use, transfer, and maintain our personal data, including limits on targeted advertising. These limits should put the burden on platforms to minimize how much information they collect, rather than burdening Americans with reading fine print. We especially need strong protections for particularly sensitive data such as geolocation and health information, including information related to reproductive health. We are encouraged to see bipartisan interest in Congress in passing legislation to protect privacy.[217]

Privacy is also among the EU's top three principles of trustworthy AI.[218] Strikingly, these principles mention no right for full and representative data collection.[219] The most pervasive privacy regulation, Europe's GDPR, presumptively prohibits all data collection or use, unless such collection is within the allowable exceptions to the privacy rule.[220]

Privacy has developed as an individual liberty: the right to be left alone. The emphasis on privacy pits two camps against one another: citizens versus the government, and individuals versus corporations. The privacy lens presents solutions to the data dilemma as binary as well: extract or conceal. We frequently overlook the costs of privacy—and the limits of these unimaginative dualities—including regressive distributional effects, innovation stagnation, and incomplete information about the root causes of inequality, poor health, insecurity, and social strife. The stakeholders left behind in the seclusion/extraction line-drawing are too often those who are at the edge of data collection—vulnerable people and communities who have not had equal access to shaping our knowledge pools.

---

215. AMERICAN DATA PRIVACY AND PROTECTION ACT DRAFT LEGISLATION, SECTION BY SECTION SUMMARY 2, https://www.commerce.senate.gov/services/files/9BA7EF5C-7554-4DF2-AD05-AD940E2B3E50 [https://perma.cc/48NY-MF5T].

216. *Id.*

217. *See* White House Briefing Room, *supra* note 12.

218. HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, ETHICS GUIDELINES FOR TRUSTWORTHY AI 14 (2019), https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html [https://perma.cc/8FRN-3T2C].

219. *Id.*

220. *See* GDPR, *supra* note 14, art. 6, at 1.

When law takes seriously the mandate of AI-for-Good, a complementary bundle of rights—co-existing, and at times competing, with privacy rights—must include a duty to collect fuller information and a corollary right to be included in data collection. For example, in health and clinical trials lies a long history of excluding women and minorities.[221] This means that the data we have now—which serves as a basis for training algorithms—is partial and skewed. Privacy serves important purposes. Yet, the notion that strong privacy rights are always beneficial and are especially protective of the vulnerable is simply not true. Indeed, when it is clear that more data will contribute to more inclusive training of AI that diminishes exclusions and bias, too much data protection may inadvertently harm vulnerable populations by deepening social inequities.

We need to recognize the existence of what Daniel Castro has called "data deserts" and "data poverty,"[222] and what Kate Crawford has described as "dark zones or shadows where some citizens and communities are overlooked or underrepresented."[223] Solon Barocas and Professor Andrew Selbst explain that those who have "unequal access to and relatively less fluency in the technology necessary to engage online, or are less profitable customers or important constituents," will have less data collected about them.[224] What we count—and what we do not count—matters immensely for resource allocation.[225] We hear a lot about the risks of creating electronic trails.[226] We hear far less about the risks of not being included in such trails. Census data, for example, as Professor Dan Bouk shows in a new book, *Democracy's Data*, has always been political.[227] Public and private entities use data all the time to decide on local investment and improvements. For instance, cities are increasingly using data from apps that detect when a mobile phone—and

221. *See, e.g.*, Patrick Boyle, *Clinical Trials Seek to Fix Their Lack of Racial Mix*, ASS'N AM. MED. COLLS. (Aug. 20, 2021), https://www.aamc.org/news/clinical-trials-seek-fix-their-lack-racial-mix [https://perma.cc/37GM-KQS6].

222. DANIEL CASTRO, CTR. FOR DATA INNOVATION, THE RISE OF DATA POVERTY IN AMERICA 2 (2014), http://www2.datainnovation.org/2014-data-poverty.pdf [https://perma.cc/GF9T-B5U4].

223. Kate Crawford, *Think Again: Big Data*, FOREIGN POL'Y (May 10, 2013, 12:40 AM), https://foreignpolicy.com/2013/05/10/think-again-big-data/ [https://perma.cc/9BG4-8PK8].

224. Barocas & Selbst, *supra* note 170, at 685. *See generally* Paul M. Schwartz, *Privacy and Participation: Personal Information and Public Sector Regulation in the United States*, 80 IOWA L. REV. 553 (1995) (comparing European and American data protection laws).

225. *See, e.g.*, Mimi Onuoha, *On Missing Datasets*, INT'L WORKSHOP ON OBFUSCATION (Oct. 5, 2017), https://www.obfuscationworkshop.org/2017/10/on-missing-datasets/ [https://perma.cc/THL9-22Y3]; Jonas Lerman, *Big Data and Its Exclusions*, 66 STAN. L. REV. ONLINE 55, 62 (2013); CATHERINE D'IGNAZIO & LAUREN F. KLEIN, DATA FEMINISM 97–98 (2020).

226. *See, e.g.*, Mary Anne Franks, *Democratic Surveillance*, 30 HARV. J. L. & TECH. 425, 452 (2017).

227. DAN BOUK, DEMOCRACY'S DATA: THE HIDDEN STORIES IN THE U.S. CENSUS AND HOW TO READ THEM (2022).

its user—has hit a pothole that requires repair and maintenance.[228] Data-rich communities will have more financial opportunities, infrastructure investment, and opportunities for civic engagement.

Privacy debates demonstrate the double-edged sword of the techlash. The contemporary stance pervasive among regulatory and civil rights activists—for example, that of the ACLU in its quest to ban the collection of biometric data—is that the loss of privacy presents an imminent threat, and that this threat is particularly acute for the most vulnerable members of society.[229] Privacy discourse marshals public/private divides, calling for freedom from state intrusion and a separation between the individual's private life and public space.[230] And yet, feminist theorists have shown that privacy has historically served men in domestic freedoms, hiding abuse from state intervention.[231] Privacy as the right to be left alone was developed by—and for—elites.[232] Professor Dan Solove describes the bundle of privacy rights as including "freedom of thought, control over one's body, solitude in one's home, control over information about oneself, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations."[233] Justice Louis D. Brandeis, the father of privacy, described privacy as "the most comprehensive of rights and the right most valued by civilized men."[234]

228. *See* Crawford, *supra* note 223 (describing the disparities cause by the Street Bump app in Boston); *see also* Sara Atske & Andrew Perrin, *Home Broadband Adoption, Computer Ownership Vary by Race, Ethnicity in the U.S.*, PEW RSCH. CTR. (July 16, 2021), https://www.pewresearch.org/fact-tank/2021/07/16/home-broadband-adoption-computer-ownership-vary-by-race-ethnicity-in-the-u-s/ [https://perma.cc/RPX9-KZGW] (highlighting the potential for racial disparity in data harvesting).

229. Press Release, ACLU, ACLU Comment on Newly Released FTC Policy Statement on Biometrics (May 19, 2023), https://www.aclu.org/press-releases/aclu-comment-on-newly-released-ftc-policy-statement-on-biometrics [https://perma.cc/4UQR-9G7F].

230. *See, e.g.*, *id.*

231. *See* Martha A. Ackelsberg & Mary Lyndon Shanley, *Privacy, Publicity, and Power: A Feminist Rethinking of the Public-Private Distinction*, *in* REVISIONING THE POLITICAL, FEMINIST RECONSTRUCTIONS OF TRADITIONAL CONCEPTS IN WESTERN POLITICAL THEORY 213, 213 (Nancy J. Hirschmann & Christine Di Stephano eds., 1996); Michele Estrin Gilman, *Welfare, Privacy, and Feminism*, 39 U. BALT. L.F. 1, 14 (2008); Suzanne A. Kim, *Reconstructing Family Privacy*, 57 HASTINGS L.J. 557, 558 (2006); CATHARINE A. MACKINNON, TOWARD A FEMINIST THEORY OF THE STATE 168–69 (1989); ANITA L. ALLEN, UNEASY ACCESS: PRIVACY FOR WOMEN IN A FREE SOCIETY 36 (1988).

232. Michele Estrin Gilman, *The Class Differential in Privacy Law*, 77 BROOK. L. REV. 1389, 1426 (2012).

233. Daniel J. Solove, *Conceptualizing Privacy*, 90 CALIF. L. REV. 1087, 1088 (2002).

234. Olmstead v. United States, 277 U.S. 438, 478 (1928) (Brandeis, J., dissenting), *overruled by* Katz v. United States, 389 U.S. 347 (1967). For more information on Justice Brandeis's role in the development of privacy law, see Leah Burrows, *To Be Let Alone: Brandeis Foresaw Privacy Problems*, BRANDEISNOW (July 24, 2013), https://www.brandeis.edu/now/2013/july/privacy.html [https://perma.cc/3TW4-YPJQ].

Privacy too often is leveraged to protect the rich and the famous from the public's right to know.[235]

2. Data Maximization

Privacy is an individual right that stands against the collective's goals. In a social democracy, we can envision subverting the script from *surveillance capitalism* to *guardianship liberalism*, imagining how, under the conditions of democratic trust, millions of surveillance cameras can become "a friendly eye in the sky, not Big Brother but a kindly and watchful uncle or aunt."[236] In a recent article, I researched the frustratingly long, stagnant reality of the gender and race pay gaps.[237] I argued that much of the problem stems from asymmetric information within the market, held by employers, putting employees and regulators at a disadvantage.[238] My research revealed that reversing information flows by publicly collecting salary data, including its gender and race breakdowns, by the Equal Employment Opportunity Commission (EEOC) as well as by privately sharing data on third-party intermediary platforms, such as Glassdoor—which has a salary calculator called *Know Your Worth*—is a way for workers to, finally, actually *know their worth*.[239] These changes are enabled by subverting cultural norms about the taboo of discussing one's salary or asking others about theirs.[240] Such information sharing is also supported by changing norms of using online platforms to crowdsource information.[241] Policy plays an active role in allowing these shifts through limiting the enforceability of certain NDAs and the definitions of company proprietary and confidential information, as well as through laws about salary disclosures and salary data collection.[242]

Law is an enabler and a blocker. Laws can make auditing for equality and nondiscrimination more difficult. Many statutes make it unlawful to ask about race or gender.[243] However, to ensure equality and prevent

---

235. *See, e.g.*, Neil M. Richards, *The Puzzle of Brandeis, Privacy, and Speech*, 63 VAND. L. REV. 1295, 1304 (2010); Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 205 (1890); ARNOW-RICHMAN ET AL., *supra* note 162, at 1; Rory Van Loo, *Privacy Pretexts*, 108 CORNELL L. REV. 1, 9–10 (2023).

236. Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 494 (2006) (quoting JEFFREY ROSEN, THE NAKED CROWD: RECLAIMING SECURITY AND FREEDOM IN AN ANXIOUS AGE 36 (2004)).

237. Lobel, *Knowledge Pays*, *supra* note 198, at 547.

238. *Id.* at 588.

239. *Id.* at 592.

240. *Id.* at 589.

241. *Id.* at 591.

242. ARNOW-RICHMAN ET AL., *supra* note 162, at 6 & 9.

243. S*ee, e.g.*, *Prohibited Employment Policies/Practices*, EEOC, https://www.eeoc.gov/prohibited-employment-policiespractices [https://perma.cc/3WA9-BNXF].

discrimination, we need that data. Indeed, there are several empirical studies suggesting that initiatives to "ban-the-box" on criminal background checks or prohibit asking about past salary may inadvertently widen gaps.[244] Whether or not such initiatives work is an empirical question that depends on the larger context in which such information is held and used. Similarly, removing identity markers from data may also create more gaps and inaccuracies. Counterintuitively, the best way to prevent discrimination may be to authorize an algorithm to collect and mine information about gender and race.[245] We generally do not have a positive mandate to collect information or to produce materials. In my current work with the Administrative Conference of the United States, for example, as a member of a consultative team on the public disclosure of agency legal materials, the focus is on what materials should be made public, not on the materials that an agency must produce.[246] Data collection is not neutral. When certain groups are underrepresented in the data used to train an algorithmic model, predictions about these groups will be inaccurate. By its very definition, a majority population has more data to be studied. A right to inclusive data collection is needed.

The GDPR contemplates many possible risks from data collection: "physical, material or non-material damage"; "discrimination, identity theft or fraud, financial loss, damage to the reputation"; "or any other significant economic or social disadvantage."[247] Those risks arising from not collecting information are not contemplated.[248] Similarly, the GDPR lists special risks that may result when "personal aspects are evaluated, in particular [analyzing] or predicting aspects concerning performance at work, economic situation, health, personal preferences or interests, reliability or [behavior], location or movements."[249] No comparative advantage principle is present to ask the question that needs to be asked: how else are people evaluated, analyzed, hired, promoted, and treated if not through judgment and decision-making based on some form of information? To state what should be obvious: there is no such thing as a decision-free or information-free world. Even in the lowest-tech of situations—a job applicant entering an office and sitting down for a face-

244. *See, e.g.*, Amanda Agan & Sonja Starr, *Ban the Box, Criminal Records, and Racial Discrimination: A Field Experiment*, 133 Q. J. ECON. 191, 195, 210 (2018).

245. *See generally* LOBEL, *supra* note 29 (discussing how algorithms can overcome bias and lead to quality in negotiating and decision-making); Talia B. Gillis, *The Input Fallacy*, 106 MINN. L. REV. 1175 (2022) (suggesting algorithmic outputs can be used to identify bias).

246. Courts interpret FOIA only to provide the public a right to see information that has been collected but does not create substantive rights about the creation of data trails. *See, e.g.*, Tax Reform Rsch. Grp. v. IRS, 419 F. Supp. 415, 418, 425 (D.D.C. 1976).

247. GDPR, *supra* note 14, at 15–16.

248. *See id.*

249. *Id.* at 15.

to-face interview—an assessment based on fact (and fiction and bias and noise) is made.[250]

The focus of privacy and data protection debates reflects an understanding that harms of exposure of information loom large: reputational harms, identity theft, user manipulation, and discrimination based on the data extracted. When a cost-benefit analysis is done balancing privacy against other values, the loss of privacy is usually weighed against efficiency, customer service, product personalization, and price precision.[251] These processes are valuable, and we should indeed consider them in the tradeoffs between stronger and weaker privacy protections. In the debates, however, the "less privacy" argument comes largely from the business economics case. It is the classic argument of "no such thing as a free lunch"—if consumers do not give up their data, they will be forced to switch to subscription models.[252] Ironically, therefore, the solutions are market-based and contractual.[253] Most of the time, the solution to the inevitable need for data in the market is that which the FTC has shaped as "notice and consent"—putting consumers or employees on notice about data collection and asking for a click on boilerplate electronic consent clauses.[254] Similarly, HIPAA

250. *See* On Amir & Orly Lobel, *Stumble, Predict, Nudge: How Behavioral Economics Informs Law and Policy*, 108 COLUM. L. REV. 2098, 2098–99 (2008) [hereinafter Amir & Lobel, *Stumble, Predict, Nudge*].

251. For example, section 104 of the American Data and Privacy Protection Act draft, titled Loyalty to Individuals with Respect to Pricing, creates certain price-based exceptions to general data protections:

> A covered entity may not retaliate against an individual for exercising any of the rights guaranteed by the Act, or any regulations promulgated under this Act, including denying goods or services, charging different prices or rates for goods or services, or providing a different level of quality of goods or services, or providing a different level of quality of goods or services. . . . Nothing in subsection (a) may be construed to . . . prohibit the relation of the price of a service or the level of service provided to an individual to the provision, by the individual, of financial information that is necessarily collected and processed only for the purpose of initiating, rendering, billing for, or collecting payment for a service or product requested by the individual; . . . [or] prohibit a covered entity from offering a different price, rate, level, quality, or selection of goods or services to an individual, including offering goods or services for no fee, if the offering is in connection with an individual's voluntary participation in a bona fide loyalty program.

American Data Privacy and Protection Act, H.R. 8152, 117th Cong. § 104(a)–(b)(2) (2022).

252. Alessandro Acquisti et al., *What is Privacy Worth?*, 42 J. LEGAL STUD. 249, 251–52 (2013).

253. And in this market setting, the findings have been termed "the privacy paradox"—consumers readily waiving their privacy rights when they recognize the benefits of free access or better service. *See id.*

254. *Id.* at 268–69.

protects patients against sharing health information without notice and consent.[255] This is also the basis of the data collection under the GDPR—platform and user agreement—an opt-in contractual regime.[256]

But what if the very fact that data is collected brings more health, safety, equality, accuracy, and socially valuable innovation? In other words, what if the tradeoffs are not simply between individual rights and cheaper services but are also between different fundamental rights? How do unequal structures of data collection and data relations mandate more, rather than less, collection of data about minorities, vulnerable populations, and women excluded, for example, from clinical trials for decades? How does data reveal lifesaving population-wide patterns and bring lifesaving innovation through completeness of the database, the speed of building accurate predictive models, and targeting collective goals, such as fighting a pandemic or building a continuous glucose monitor and closed-loop insulin pump? As Professor Pedro Domingos put it, "More Data Beats a Cleverer Algorithm."[257]

The GDPR prohibits the processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, and genetic or biometric data for the purpose of uniquely identifying a person.[258] Data collection concerning health, a person's sex life, or sexual orientation is also prohibited under the GDPR.[259] There are important exceptions when the processing is necessary for the

> purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems . . . processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices.[260]

The processing exception still must be "proportionate to the aim pursued, respect the essence of the right to data protection[,] and provide

255. Health Insurance Portability and Accountability Act of 1996 (HIPAA), Pub. L. No. 104-191, 110 Stat. 1936, 2033 (codified as amended in scattered sections of 18, 26, 29, and 42 U.S.C.); 45 CFR §§ 164.506(b), 164.520 (2013).

256. GDPR, *supra* note 14, at 6, 36.

257. Pedro Domingos, *A Few Useful Things to Know About Machine Learning*, 55 COMMC'NS OF THE ACM 78, 84 (2012). It is also worth noting that AI can help mitigate some of these tensions between the need for data and privacy, for example, by producing synthetic data. *See* Peter Lee*, Synthetic Data & The Future of AI* (unpublished manuscript) (on file with author).

258. GDPR, *supra* note 14, at 38.

259. *Id.*

260. *Id.*

for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject."[261] These exceptions are narrow and primarily focus on health-related and research purposes.[262] However, there are many more valuable social goals that benefit from data maximization: preventing violence, trafficking, and hate crimes; increasing equality, diversity and inclusion in workplaces, products, and service markets; promoting road safety and fuel efficiency; and improving access and learning in education school systems.

In Europe, the application of the GDPR's broad regulatory framework will vary across member states and will depend on the details of operationalizing each rule in particular contexts. The exceptions may be construed widely enough to include many positive data collection purposes. However, both in Europe and in the United States, setting defaults that privilege privacy combined with a techlash policy mindset makes such a balanced construction unlikely. Even when it comes to one of the most clearly acceptable exceptions to privacy, that of scientific research, the GDPR has already begun to present challenges.[263] For market innovation and competition, an unbalanced application of the GDPR and any imbalanced privacy law may result in a range of unintended regressive effects.[264]

C. *Frontiers of Proactive AI Policy*

1. Public-Private AI Governance

AI presents new opportunities to reach a desirable—yet delicate—balance. Technology requires regulation and opens opportunities for more ways to regulate. I have long argued against a false dilemma between centralized command-and-control regulation and collaborative private-public governance.[265] As a matter of public policy, we should

261. *Id.* at 39.

262. *Id.*

263. *See* EUR. SOC'Y OF HUM. GENETICS, *Balancing Data Protection and Research Needs in the Age of the GDPR*, SCIENCEDAILY (June 17, 2019), https://www.sciencedaily.com/releases/2019/06/190617100942.htm [https://perma.cc/UF9Q-LG9C].

264. *See* Tal Z. Zarsky, *Incompatible: The GDPR in the Age of Big Data*, 47 SETON HALL L. REV. 995, 1020 (2017). *See generally* Joel Thayer, *Can a Machine Learn Under the GDPR?*, TPRC 46: THE 46TH RSCH. CONF. ON COMM'N, INFO., & INTERNET POL'Y 2918 (Dec. 16, 2018), https://ssrn.com/abstract=3141854 [https://perma.cc/BA2L-ZFGD] (discussing the uncertainty of the GDPR's effect on machine learning).

265. *See, e.g.*, Orly Lobel, *The Renew Deal: The Fall of Regulation and the Rise of Governance in Contemporary Legal Thought*, 89 MINN. L. REV. 342, 343 (2004); Lobel, *Interlocking Regulatory and Industrial Relations*, *supra* note 150, at 1072; Orly Lobel, *New Governance as Regulatory Governance*, *in* OXFORD HANDBOOK OF GOVERNANCE 65, 65 (David Levi-Faur ed., 2012); Orly Lobel & On Amir, *Liberalism and Lifestyle: Informing Regulatory Governance with Behavioral Research*, 1 EUR. J. RISK REG. 17, 17 (2012) [hereinafter Lobel & Amir, *Liberalism and Lifestyle*].

invest in better AI both for infrastructure–enabling massive and secured collection of data that is automatically generated and managed—and for decision-making processes—mining data, detecting patterns, and identifying courses of action.

Pay equity, which we examined above, demonstrates the potential for shared governance, that is, relying on public data collection and private platform information sharing.[266] There are promising developments in this direction. On September 14, 2022, the New York Stock Exchange announced a new partnership with Syndio, a leading software company that provides businesses with AI tools to analyze, resolve, and prevent gender and race disparities in screening, hiring, and compensation.[267] The SEC proposed rules to enhance disclosures regarding corporate diversity, and the House Financial Services Committee released draft legislation for the Ensuring Pay Equity Act, which would require financial agencies covered under Dodd-Frank to conduct internal pay equity audits.[268] The EEOC's Strategic Plan for 2022 indicates new priorities around "[c]ollecting pay data" and enforcing pay equity.[269]

These private and public initiatives require attention to best practices in data collection and AI deployment. Government agencies need to proactively support industries in sorting the best systems of software as a service–AI systems provided in the market that help companies analyze complete, real-time data.[270] Such AI-driven audits move beyond the limitations of one-time or annual audits. As new technologies support smarter governance, government agencies must also become research and

266. *See generally* Lobel, *Knowledge Pays*, *supra* note 198 (discussing the benefit public data collection has on pay equity).

267. Katherine Doherty, *NYSE Adds Pay Equity, Opportunity Tracking Tools in ESG Push*, BLOOMBERG (Sept. 13, 2022, 8:15 AM), https://www.bloomberg.com/news/articles/2022-09-13/nyse-adds-pay-equity-opportunity-tracking-tools-amid-esg-push [https://perma.cc/YJN7-DXKE].

268. Press Release, Sec. & Exch. Comm'n, SEC Proposes to Enhance Disclosures by Certain Investment Advisers and Investment Companies about ESG Investment Practices (May 25, 2022), https://www.sec.gov/news/press-release/2022-92 [https://perma.cc/C2SQ-QEKZ]. On October 27, 2021, the House Financial Services Committee hosted a hearing in which draft legislation for the "Ensuring Pay Equity Act" was published and discussed. *Bringing Consumer Protection Back: A Semi-Annual Review of the Consumer Financial Protection Bureau: Hybrid Hearing Before H. Fin. Servs. Comm.*, 117th Cong. (2021), https://financialservices.house.gov/calendar/eventsingle.aspx?EventID=408175 [https://perma.cc/6DK2-QVSZ]; H. FIN. SERVS. COMM., PAY EQUITY ACT DISCUSSION DRAFT, https://democrats-financialservices.house.gov/uploadedfiles/10.27_bills-117pih-theensuringpayequityact.pdf [https://perma.cc/L6ZS-JP9S].

269. Press Release, U.S. Equal Emp. Opportunity Comm'n, EEOC Announces Independent Study Confirming Pay Data Collection is a Key Tool to Fight Discrimination (July 28, 2022), https://www.eeoc.gov/newsroom/eeoc-announces-independent-study-confirming-pay-data-collection-key-tool-fight [https://perma.cc/YU37-AEMX].

270. *See Is Your Company Ready for Pay Equity and DEI in 2022?*, SYNDIO (Dec. 6, 2021), https://synd.io/blog_post/2022-promises-to-be-an-eventful-year-for-pay-equity-and-dei-is-your-company-ready/ [https://perma.cc/LB4G-C62B].

development arms that incentivize, test, approve, and monitor private innovation. New regulatory tools such as innovation sandboxes,[271] tech procurement,[272] and testbeds—formal settings for conducting rigorous evaluations of AI[273]—are important directions to support AI-for-Good proactive policies.

In August 2022, President Joe Biden signed into law the biggest investment package for American science and technology in many years.[274] This is a highly promising development. The CHIPS and Science Act is designed to support STEM careers and rebuild the United States' dominance in science, technology, and innovation.[275] One aspect of the Act is authorizing the National Institute for Science and Technology (NIST) to establish artificial intelligence and machine learning testbeds.[276] This investment in AI testbeds is unique among a sea of other federal tech policies. Similar to testbeds, regulatory sandboxes, as defined by the European Council in 2020, are

> concrete frameworks which, by providing a structured context for experimentation, enable where appropriate in a real-world environment the testing of innovative

271. *See, e.g.*, *The Lawtech Sandbox*, TECH NATION, https://technation.io/lawtech-sandbox/ [https://perma.cc/FQ8T-UM23] (describing a "sandbox" as an exploratory space that "fast tracks" innovation); Chang-Hsien Tsai et al., *The Diffusion of the Sandbox Approach to Disruptive Innovation and Its Limitations*, 53 CORNELL INT'L L.J. 261, 261 (2020) (discussing regulatory sandboxes); Si Ying Tan & Araz Taeihagh, *Adaptive Governance of Autonomous Vehicles: Accelerating the Adoption of Disruptive Technologies in Singapore*, 38 GOV'T INFO. Q. 101545, 101551 (2021) (explaining regulatory sandboxes in Singapore's autonomous vehicle market); Sofia Ranchordas, *Experimental Regulations for AI: Sandboxes for Morals and Mores* 17–19 (Univ. of Groningen Fac. of L. Rsch., Working Paper No. 7, 2021), https://ssrn.com/abstract=3839744 [https://perma.cc/422K-4MNN] ("AI regulatory sandboxes . . . [established by one or more] Member States competent authorities or the European Data Protection Supervisor . . . 'shall provide a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market or putting into service pursuant to a specific plan.' [This shall take place under the direct supervision and guidance] by the competent authorities 'with a view to ensuring compliance with the requirements of this Regulation and, where relevant, other Union and Member States legislation supervised within the sandbox.").

272. *See, e.g.*, Cary Coglianese & Erik Lampmann, *Contracting for Algorithmic Accountability*, 6 ADMIN. L. REV. ACCORD 175, 179–80 (2021).

273. *See, e.g.*, Tina Huang, *Creating an AI Testbed for Government*, DAY ONE PROJECT. Jan. 19, 2022, https://progress.institute/creating-an-ai-testbed-for-government/ [https://perma.cc/Q62X-VVZT].

274. *See* Press Release, WHITE HOUSE BRIEFING ROOM, Fact Sheet: CHIPS and Science Act Will Lower Costs, Create Jobs, Strengthen Supply Chains, and Counter China (Aug. 9, 2022), https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/ [https://perma.cc/G65J-RTJD].

275. CHIPS and Science Act of 2022, Pub. L. No. 117-167, 136 Stat. 1366, 1510 (codified as amended in scattered sections of 15 U.S.C.).

276. *Id.* at § 10232.

> technologies, products, services or approaches . . . for a limited time and in a limited part of a sector or area under regulatory supervision ensuring that appropriate safeguards are in place.[277]

Testbeds and sandboxes are examples of frameworks that shift the role of policy from reactive and adaptive to proactive and anticipatory.[278]

2. Bug Bounties, Sandboxes, and Testbeds

Another example of proactive AI governance is creating public bias bounties. Bias bounties are systems modeled after bug bounty systems. Bug bounties, prevalently used by the tech industry and the Department of Defense, are reward systems that outsource the findings of hacking and cybersecurity vulnerabilities to third parties, such as private and nonprofit platforms such as BugCrowd and HackerOne, which globally crowdsource the hacking efforts.[279] A 2022 report by the Algorithmic Justice League sets forth principles on how to extend bug bounty programs to social issues.[280] Governments can also support the development of and access to auditing tools that detect bias and open-source computational efforts.[281] Standardization is an important

277. Council Conclusions on Regulatory Sandboxes and Experimentation Clauses as Tools for an Innovation-Friendly, Future-Proof and Resilient Regulatory Framework That Masters Disruptive Challenges in the Digital Age, 2020 O.J. (C 447) 1, 2.

278. Public agencies can create AI competitions. Research already provides such a competitive model. For example, a study looking at nearly two thousand machine-learning algorithms used to predict breast cancer risk recently revealed one algorithm as the most accurate of the existing technology. Ricvan Dana Nindrea et al., *Diagnostic Accuracy of Different Machine Learning Algorithms for Breast Cancer Risk Calculation: a Meta-Analysis*, 19 ASIAN PAC. J. OF CANCER PREVENTION 1747, 1747 (2018). *See generally* Iqbal H. Sarker, *Machine Learning: Algorithms, Real-World Applications and Research Directions*, 2 SN COMPUT. SCI. 159 (2021), https://doi.org/10.1007/s42979-021-00592-x [https://perma.cc/W4JT-M3T2] (explaining effective AI research methods and providing real-world examples); Antonio A. Ginart et al., C*ompeting AI: How does competition feedback affect machine learning?*, 130 PROC. OF MACH. LEARNING RSCH., 2021, https://arxiv.org/pdf/2009.06797.pdf [https://perma.cc/3GZW-8LN9] (arguing that competition is necessary for the beneficial development of AI).

279. *See* JOSH KENWAY ET AL., ALGORITHMIC JUST. LEAGUE, BUG BOUNTIES FOR ALGORITHMIC HARMS? 7 (2022), https://www.ajl.org/bugs [https://perma.cc/PX3D-WA3B].

280. *Id.*; *see also* Rumman Chowdhury & Jutta Williams, *Introducing Twitter's First Algorithmic Bias Bounty Challenge*, TWITTER ENG'G BLOG (July 30, 2021), https://blog.twitter.com/engineering/en_us/topics/insights/2021/algorithmic-bias-bounty-challenge [https://perma.cc/D5SS-CUXU] (announcing the challenge used as the case study in the Algorithmic Justice League report); Khari Johnson, *AI Researchers Propose 'Bias Bounties' to Put Ethics Principles into Practice*, VENTUREBEAT (Apr. 17, 2020, 8:05 AM), https://venturebeat.com/2020/04/17/ai-researchers-propose-bias-bounties-to-put-ethics-principles-into-practice/ [https://perma.cc/3KDS-QQ39] (arguing in favor of bias bounties).

281. At the same time, we need to recognize tradeoffs between open shared AI and system vulnerability. One of the great concerns now with AI deployment is that AI can be vulnerable to

complement to such initiatives. Researchers have proposed "model cards for models" and "datasheets for datasets" reporting the use of AI—short documents that will accompany algorithms to disclose how the model performs across demographic groups that create consistency for comparison, audits, and learning.[282] In 2021, the NIST released a proposal calling on the tech community to develop voluntary, consensus-based standards for detecting AI bias, including examining, detecting, and monitoring for biases during all stages of an AI lifecycle—planning and conceptualization, design, and usage.[283] In 2020, Congress enlisted the NIST to develop a framework for managing risks of AI systems.[284] The agency has released a Draft Concept Paper which has the most balanced language I have seen in recent policy drafts.[285] It explains:

> While some interpretations of consequence focus exclusively on negative impacts (what is the likelihood that something bad will happen?), NIST intends to use a broader definition that offers a more comprehensive view of potential influences, including those that are positive, resonating with the goals of developing and applying AI technologies to achieve positive outcomes.[286]

---

malicious tactics and security breaches. This includes "data poisoning," where an adversary (or simply bad actors hacking into systems) tamper with the data environments, causing an AI to learn from inaccurate data. Paddy Smith, *Data Poisoning: A New Front in the AI Cyber War*, AI MAG. (Oct. 8, 2020), https://aimagazine.com/data-and-analytics/data-poisoning-new-front-ai-cyber-war [https://perma.cc/62WV-JKPW]. This in turn means that policy must also grapple with transparency tradeoffs. Related to this dilemma is that of data and automation capability sharing, increasing competition versus concentration and secrecy. *See* Claire Leibowicz et al., *How to Share the Tools to Spot Deepfakes (Without Breaking Them)*, MEDIUM (Jan. 13, 2022), https://medium.com/partnership-on-ai/how-to-share-the-tools-to-spot-deepfakes-without-breaking-them-53d45cd615ac [https://perma.cc/35TB-JVKJ]; *see also* Elena Chachko, *National Security by Platform*, 25 STAN. TECH. L. REV. 55, 140 (2021) (reporting on the regulation of terrorist and violent content online).

282. *See* Margaret Mitchell et al., *Model Cards for Model Reporting*, FAT '19: PROC. CONF. ON FAIRNESS, ACCOUNTABILITY & TRANSPARENCY 220, 220 (2019), https://doi.org/10.1145/3287560.3287596 [https://perma.cc/MU77-2VF9] (suggesting model cards as a supplement to datasheets); Timnit Gebru et al., *Datasheets for Datasets*, 64 COMMC'NS OF THE ACM 86 (2021), http://dx.doi.org/10.1145/3458723 [https://perma.cc/P762-YQF2] (recommending the development for datasheets).

283. *AI Risk Management Framework Concept Paper*, NAT'L INST. OF STANDARDS & TECH. (Dec. 13, 2021), https://www.nist.gov/document/airmfconceptpaper [https://perma.cc/5NUC-HWAA].

284. H.R. Rep. No. 116-455, at 23 (2021).

285. *NIST Seeks Comments on Concept Paper for AI Risk Management Framework*, NAT'L INST. OF STANDARDS & TECH. (Dec. 14, 2021), https://www.nist.gov/news-events/news/2021/12/nist-seeks-comments-concept-paper-ai-risk-management-framework [https://perma.cc/FF35-3HDL].

286. *AI Risk Management Framework Concept Paper*, *supra* note 283.

The initial draft reiterates the rejection of a one-side AI-as-Wrongs approach: "While risk management processes address adverse impacts, this framework intends to offer approaches to minimize anticipated negative impacts of AI systems and identify opportunities to maximize positive impacts."[287]

A final, highly critical proactive path for public policy, which we now turn to, is implementing behavioral lessons about human-machine interaction. Public policy needs to ask how technology can support our shared goals; our physical, cognitive, and emotional needs; and our inevitable and ongoing fallibilities. The 2023 President's Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence issued on October 30, 2023 charges several federal agencies, including NIST, with developing guidelines and best practices on the development and use of AI.[288] The Executive Order directs NIST among other things to "launch a new initiative to create guidance and benchmarks for evaluating and auditing AI capabilities, with a focus on capabilities that could cause harm."[289] It further directs NIST to help ensure the availability of testing environments, such as testbeds, to support the development of safe, secure, and trustworthy AI technologies.[290] While most of the executive order continues the focus on AI harms, the calls for best practices and standards is an important step toward a law of AI for good.

D. *Behavioral Law of AI Trust (Debiasing Humans re: Algorithms)*

1. Between Algorithmic Aversion and Algorithmic Adoration

The adoption of AI is bound to accelerate, affecting every aspect of our lives. At the same time, contemporary tech policy scholarship, public debates, and reform proposals pervasively question automation as a desirable development. As a community, we thus have a heightened awareness that there are harms and risks associated with automation. But we do not yet have a common language, or even shared taxonomy, to compare and evaluate the tradeoffs inherent to automation. I call this the "human-AI trust gap," which I argue is a significant barrier to benefiting from automation opportunities. That is, whether we have too little or too much trust in algorithms, the human-AI trust gap is that we are missing a shared literature and methods to understand when trust is given and when trust is due. Government entities should commit to improving AI and building rational social trust in these systems. Policymakers must study

287. *AI Risk Management Framework: Initial Draft*, NAT'L INST. OF STANDARDS & TECH. (Mar. 17, 2022), https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf [https://perma.cc/WBL8-UDGZ].

288. Biden Executive Order, *supra* note 7.

289. *Id.*

290. *Id.*

how to effectively integrate AI tools within human processes and systems. Digital literacy—and improving digital rationality—should be a national strategy.[291] The aim should be the right mix of trust and skepticism—a Goldilocks appreciation of technology based on accurate assessments and acceptable trade-offs.

Behavioral human-AI research, which examines algorithmic trust and human-algorithm interactions, is a rather nascent field of study. Like the overarching field of behavioral studies, much of the insights come from the business schools, particularly the marketing literature, which tends to focus on how consumers make decisions.[292] For example, scholarly experiments and private digital platforms invest in understanding

291. The United Kingdom, New Zealand, South Korea, and other countries are making strides on this. *See, e.g.*, BENSON NEETHIPUDI ET AL., CTR. FOR UNIVERSAL EDUC. AT BROOKINGS, HOW SOUTH KOREA IMPLEMENTED ITS COMPUTER SCIENCE EDUCATION PROGRAM 2 (2021), https://www.brookings.edu/wp-content/uploads/2021/10/How-S-Korea-implemented-its-CS-program_FINAL.pdf [https://perma.cc/L2SM-S7AT]; Aineena Hani, *Implementing Digital Technology Learning in New Zealand Schools*, OPENGOV ASIA (July 9, 2021), https://opengovasia.com/implementing-digital-technology-learning-in-new-zealand-schools/ [https://perma.cc/S2E7-J2RY]; Julian McDougall, *A New Media Literacy Education Bill?*, LONDON SCH. OF ECON. & POL. SCI. (June 27, 2022), https://blogs.lse.ac.uk/medialse/2022/ 06/27/a-new-media-literacy-education-bill/ [https://perma.cc/73GC-RC29]. The UK All-Party Parliamentary Group on Media Literacy published a report that recommended a Media Literacy Education Bill. *See* U.K. DEP'T FOR DIGIT., CULTURE, MEDIA & SPORT, ONLINE MEDIA LITERACY STRATEGY 68 (2021), https://assets.publishing.service.gov.uk/government/uploads/system/up loads/attachment_data/file/1004233/DCMS_Media_Literacy_Report_Roll_Out_Accessible_PDF.pdf [https://perma.cc/KBF8-34EW] (describing how the UK "government is carrying out research to help boost UK citizens' resilience to dis- and misinformation in the UK"). The importance of industrial policy for AI is that, as we saw, the private sector is forging ahead while regulators are almost exclusively focused on creating safeguards against AI risks. Government intervention in markets to promote the development of useful tech is critical to the success of the U.S. economy. *See* Todd N. Tucker & Steph Sterling, *Industrial Policy and Planning: A New (Old) Approach to Policymaking for a New Era*, ROOSEVELT INST. 5 (Aug. 2021), https://rooseveltinstitute.org/wp-content/uploads/2021/07/RI_ANewOldApproachtoPolicymakingforaNewEra_IssueBrief_202108.pdf [https://perma.cc/6Q4Y-FBUG]; *The U.S. Innovation and Competition Act: Senate Passes Sweeping $250 Billion Bill to Bolster Scientific Innovation and Compete With China*, SIDLEY: GOV'T STRATEGIES UPDATE (June 16, 2021), https://www.sidley.com/en/insights/newsupdates/2021/06/an-overview-of-the-united-states-innovation-and-competition-act [https://perma.cc/K98E-TC5Q]. *See generally* Orly Lobel, *Biopolitical Opportunities: Between Datafication and Governance*, 96 NOTRE DAME L. REV. REFLECTION 181 (2021) [hereinafter Lobel, *Biopolitical Opportunities*] (considering ways in which governments can engage in new forms of governance to leverage the biopolitical data extracted by private actors for profit in service of public goals of fairness, equality, and distributive justice).

292. *See, e.g.*, Amir & Lobel, *Stumble, Predict, Nudge*, *supra* note 250, at 2099; On Amir & Orly Lobel, *Risk Management for the Future: Age, Risk and Choice Architecture*, 43 ADVANCES IN CONSUMER RSCH. 93, 93 (2015); Lobel & Amir, *Liberalism and Lifestyle*, *supra* note 265. There is also nascent related literature on how the very concept of trust, including in our democratic processes, is changing with new technology. *See, e.g.*, M. TODD HENDERSON & SALEN CHURI, THE TRUST REVOLUTION 125 (2019).

algorithmic trust as it relates to automated recommenders.[293] Similar research needs to be done at the policy level. For decades, the policy implications of behavioral studies have lagged behind market applications, and I predict that, similarly, we will soon see a more concerted effort to understand the policy implications of behavioral human-AI studies.

Algorithmic trust—and distrust—is multidimensional. A 2022 Pew survey, consistent with other recent studies, finds that most Americans fear AI and have little confidence about its use by government entities.[294] Ironically, we fear AI's flaws and flawlessness. On the one hand, critics point to the failures of AI and its "rudimentary" state, recalling the FTC report explored above.[295] Simultaneously, critics lament its uber-rationality and consistency: "[M]aking decisions via the rigid, rule-based logic of algorithms violates the principle that government decisions should respond to the circumstances of individual people."[296] Studies also find demographic differences in AI trust. For example, women view AI more negatively than men.[297] Education and income levels are also predictors of AI aversion, with lower education and income correlating with higher distrust.[298] As I suggested earlier, there is likely a generational difference in the willingness to give up control and allow a

293. *See, e.g.*, Johannes Kunkel et al., *Let Me Explain: Impact of Personal and Impersonal Explanations on Trust in Recommender Systems*, *in* CHI '19: PROC. OF THE 2019 CHI CONF. ON HUM. FACTORS IN COMPUT. SYS., PAPER NO. 487, at 1, 2 (May 4, 2019), https://doi.org/10.1145/3290605.3300717 [https://perma.cc/VX2E-F7M6]; Michael Jugovac & Dietmar Jannach, *Interacting with Recommenders—Overview and Research Directions*, 7 ACM TRANSACTIONS ON INTERACTIVE INTEL. SYS. no. 3, 2017, at 1 (2017), https://doi.org/10.1145/3001837 [https://perma.cc/BZL7-2YB8].

294. LEE RAINIE ET AL., AI AND HUMAN ENHANCEMENT: AMERICANS' OPENNESS IS TEMPERED BY A RANGE OF CONCERNS 20, 23 (2022), https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2022/03/PS_2022.03.17_AI-HE_REPORT.pdf [https://perma.cc/DSM4-KVAF]; BAOBAO ZHANG & ALLAN DAFOE, ARTIFICIAL INTELLIGENCE: AMERICAN ATTITUDES AND TRENDS 20 (2019), https://isps.yale.edu/sites/default/files/files/Zhang_us_public_opinion_report_jan_2019.pdf [https://perma.cc/3LXF-6VJ5].

295. *See* FED. TRADE COMM'N, *supra* note 31, at 75.

296. Ben Green, *The Flaws of Policies Requiring Human Oversight of Government Algorithms*, 45 COMPUT. L. & SEC. REV., no. 105681, 2022, at 1.

297. *See, e.g.*, EUROPEAN COMM'N, DIRECTORATE-GENERAL FOR COMMUNIC'N, SPECIAL EUROBAROMETER 460 REPORT ON ATTITUDES TOWARD THE IMPACT OF DIGITISATION AND AUTOMATION ON DAILY LIFE 60 (May 2017), https://data.europa.eu/doi/10.2759/835661 [https://perma.cc/UC83-SJX8]; MORNING CONSULT, NATIONAL TRACKING POLL #170401, at 34 (2017), https://morningconsult.com/wp-content/uploads/2017/04/170401_crosstabs_Brands_v3_AG.pdf [https://perma.cc/3HMQ-EXU6].

298. *See, e.g.*, AARON SMITH & MONICA ANDERSON, AUTOMATION IN EVERYDAY LIFE 9 (2017), https://www.pewresearch.org/internet/2017/10/04/automation-in-everyday-life/ [https://perma.cc/4ZG2-VKL6]; *cf.* Alex J. Wood et al., *Good Gig, Bad Gig: Autonomy and Algorithmic Control in the Global Gig Economy*, 33 WORK, EMP., & SOC'Y 56, 69–70 (2019) (discussing the advent of the "gig economy" and its socioeconomic impact).

robot to do for your children what your parents had done for you, such as driving.[299]

Under certain circumstances, we trust bots too much. We might call this "algorithmic adoration"—some behavioral scientists have termed it "algorithm appreciation," but that does not seem to capture an over-trust attitude.[300] We have long held ambivalent, and even irrational, attitudes toward technology, and in some studies, humans are found to be too trusting and perceive algorithms as inherently superior to human decision-makers.[301] The technical nature of AI tools may convey a false sense of precision and objectivity, lending a sense of inevitability to outcomes that in fact rest on human choices. In one experiment at Georgia Tech, participants were so ready to trust a robot that they were willing to follow it toward what seemed to be a burning building, using pathways that were clearly wrong and inconvenient.[302] Ironically, however, other experiments find that people become less trusting of bots after realizing that bots outperform humans.[303] Part of the responsibility of government regulators is to understand why and when people are averse to algorithms or inherently prefer a human decision-maker. Moreover, educational efforts can help moderate the irrationalities of both algorithm aversion and algorithm adoration.

It may be that some of the psychology of algorithmic aversion is a particular iteration to more general behavioral traits. People tend to accept known risks, while fearing risks that are new, unknown, and not well understood. As I described at the outset of this Article, we tend to prefer the status quo (e.g., status-quo bias) and be more averse to losses compared to feeling the pain of forgone potential gains.[304]

It is worth mentioning the views of two Nobel laureate behavioralists, Daniel Kahneman and Richard Thaler, on human-AI trust. In a 2021 interview, building on his vast body of research and his two recent popular books, Kahneman lamented that people are still far more inclined

299. London, *supra* note 187.

300. *See generally* Jennifer M. Logg et al., *Algorithm Appreciation: People Prefer Algorithmic to Human Judgment*, 151 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 90 (2019) (finding that people gave more weight to advice from algorithmic rather than human sources); Raja Parasuraman & Dietrich H. Manzey, *Complacency and Bias in Human Use of Automation: An Attentional Integration*, 52 HUM. FACTORS 381 (2010) (contending that users may misuse automated decision aids due to a tendency to place more trust in automated aids than other sources of advice).

301. Parasuraman & Manzey, *supra* note 300, at 391.

302. Paul Robinette et al., *Overtrust of Robots in Emergency Evacuation Scenarios*, *in* HRI '16: 11TH ACM/IEEE INT'L CONF. ON HUMAN-ROBOT INTERACTION 101, 106 (Mar. 7, 2016).

303. *See, e.g.*, Berkeley J. Dietvorst et al., *Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err*, 144 J. EXPERIMENTAL PSYCH.: GEN. 114, 115 (2015).

304. *See supra* note 22 and accompanying text.

to trust human systems than artificial ones.[305] He gives an example analogous to the aversion of AI—that of irrational aversion to vaccination:

> We see that, for example, in terms of the attitude to vaccination. People are willing to take far, far fewer risks when they face vaccination than when they face the disease. So this gap between the natural and the artificial is found everywhere. In part that is because when artificial intelligence makes a mistake, that mistake looks completely foolish to humans, or almost evil.[306]

Richard Thaler recently described his understanding of AI with "two one-liners that happened to have been uttered by brilliant Israelis."[307] Thaler described how social psychologist Amos Tversky, when asked about AI, responded that "he did not know much about it, his specialty [was] natural stupidity."[308] Thaler also describes how Israeli Ambassador Abba Eban was asked if Israel would switch from a six-day to a five-day workweek: "Eban's reply to the query about a five-day workweek was: 'One step at a time. First, let's start with four days, and go from there.'"[309] The import of both anecdotes: humans are deeply—and wonderfully—imperfect. Thaler confesses to finding "the stubborn reluctance in many segments of society to allow computers to take over tasks that simple models perform demonstrably better than humans"[310] far more alarming than any distant possibility of AI running amok.

---

305. Adams, *supra* note 188.

306. *Id.* Indeed, we see this attitude of aversion toward "artificial" human-made creations as opposed to "natural" ones in many other irrational judgements, and we know that there must be a role for policy, by design and education, to offset such irrationalities. *Id.* For example, fears of GMO—a technology that can help alleviate food hunger around the world—overlook the fact that everything we eat today has been genetically modified for centuries, since our ancestors mutated vegetable variants to make them more durable. Alicia Hills Moore, *Monsanto's Bet: There's Gold in Going Green CEO Robert Shapiro Thinks He Can Feed the World's Exploding Population and Heal the Environment with Genetically Engineered Crops. He also Thinks He Can Make a Lot of Money for Shareholders Along the Way.*, CNN MONEY (Apr. 14, 1997), https://money.cnn.com/magazines/fortune/fortune_archive/1997/04/14/224981/ [https://perma.cc/X8T3-HGHF]. The ancient Greeks grafted plants to the roots of other plants. The life-saving human-manufactured insulin is the first genetically engineered drug. Michael Specer, *The Pharmageddon Riddle*, THE NEW YORKER, Apr. 2, 2000, at 58. I thank Miranda Fleisher and Robert Shapiro for the conversations about GMOs. A federal bill, H.R. 1599—the Safe and Accurate Food Labeling Act of 2015—passed the House but died in the Senate. H.R. 1599, 114th Cong. It would have adopted a national standard for labeling laws related to GMOs and prohibited a requirement imposed on food companies to disclose their use of genetically modified ingredients. *See id.* § 103(f).

307. Richard H. Thaler, *Who's Afraid of Artificial Intelligence?*, EDGE (2015), https://www.edge.org/response-detail/26083 [https://perma.cc/L6XN-ZY6F].

308. *Id.*

309. *Id.*

310. *Id.*

2. Inadvertent Irrationality in Contemporary Policy

Whichever keeps us up at night, human or AI fallibility, we need to devote more public attention to human-machine interaction. There is already a body of research that can give us clues to possible policy directions, including questioning some of the recent AI policy proposals currently underway. For one, does knowing that the decision-maker is artificial help? The right to know that you are interacting with a bot, or that you are subject to automated decision-making, is a centerpiece of EU/U.S. legislative proposals. For example, under the EU Draft AI Act, consumers would have a right to see disclosures that they are chatting with or seeing images produced by AI.[311] In 2021, Quebec similarly passed a law that requires individuals to be informed when automated decision-making tools are being used.[312] Yet, research shows that this right to know about automation may have inadvertent harms.

In a recent experiment published in *Nature*, physicians received chest X-rays and diagnostic advice, some of which were inaccurate.[313] While the advice was all generated by humans, some of the advice for the purpose of the experiment was labeled as generated by AI and some by human experts.[314] In the experiment, radiologists rated the same advice as lower quality when it appeared to come from an AI system.[315] Other studies find that, when the recommendations pertain to more subjective types of decisions, humans are even less likely to rely on the algorithm.[316] This holds true even when the subjects see the algorithm outperform the human and when they witness the human make the same error as the algorithm.[317] At the same time, in education, researchers have found that when a robot mimics human fallibility, the child's learning process supported by the AI improves.[318]

In the context of air travel, in a new research project with On Amir, Paul Wynns, and Alon Pereg, my collaborators and I seek to design a

311. EU Draft AI Act, *supra* note 15, at 69 ("Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use.").

312. Samuel Adams, *Quebec's Bill 64: The First of Many Privacy Modernization Bills in Canada?*, INT'L ASS'N PRIV. PRO. (Nov. 23, 2021), https://iapp.org/news/a/quebecs-bill-64-the-first-of-many-privacy-modernization-bills-in-canada/ [https://perma.cc/V8HC-E8FN].

313. Susanne Gaube et al., *Do as AI Say: Susceptibility in Deployment of Clinical Decision-Aids*, 4 NATURE PARTNER J. DIGIT. MED., 2021, at 2.

314. *Id.*

315. *Id.*

316. Berkeley J. Dietvorst & Soaham Bharti, *People Reject Algorithms in Uncertain Decision Domains Because They Have Diminishing Sensitivity to Forecasting Error*, 31 PSYCH. SCI. 1302, 1310–11 (2020).

317. *Id.*

318. LOBEL, *supra* note 29.

series of experiments to examine what factors contribute to trust or distrust in automation, even in situations where pilots know automation is superior. This work anticipates the next phase in aviation where completely autonomous vehicles will transport people. It also builds on initial evidence, that, notwithstanding my optimism about algorithmic trust given the prevalence of autopilot use, most passengers are not entirely aware of this prevalence.

The sphere of automation seems to matter significantly. In conversations with colleagues and policymakers, I have encountered a repeated fallacy—call it the "Bot! stay in your lane" fallacy—where even when we admit some human tasks can be better done by AI, there is a desire to confine that sphere to a neatly defined set of capabilities: computational, acontextual, quantifiable, and objective. These are absolutely domains where AI has developed faster than others, yet recent advances like Open AI's DALL-E and GPT are proving that AI is getting better in many domains that we would associate with contextual, creative, subjective, emotional, and moral reasoning.[319]

Relatedly, a study that examined the effects of the replacement of a human by an automated system showed that, over the course of twenty forecasting trials, people trusted automated advisors less if the automated advisor had replaced an initial human one, as opposed to having been introduced from the beginning of the forecasting interactions.[320] Additionally, automated advisors that replaced humans were rated as issuing lower quality advice, while human advisors that replaced automated advisors were rated as providing better quality advice.[321] In addition to seeing an algorithm make a mistake or issue poor advice, algorithm aversion also increased when people had to choose between an algorithm's forecast and their own, particularly when the people choosing had expertise in the subject they were forecasting.[322]

Does explanation improve trust? The right to explainable AI is part of EU and American AI regulatory reforms.[323] But behavioral research

319. For more on this divide and subjective decisions, see *infra* note 328 and accompanying text.

320. Andrew Prahl, *Algorithm Admonishment: People Distrust Automation More When Automation Replaces Humans* 20–24 (Working Paper, Jan. 8, 2020), http://dx.doi.org/10.2139/ssrn.3903847 [https://perma.cc/CGJ6-5JT3].

321. *Id.* at 23.

322. *See* Logg et al., *supra* note 300.

323. Gianclaudio Malgieri, *Automated Decision-Making in the EU Member States: The Right to Explanation and Other "Suitable Safeguards" in the National Legislations*, 35 COMPUT. L. & SEC. REV., 2019, at 10; Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 192 (2019). The CPRA's automated decision-making provisions require businesses to provide a consumer with "meaningful information about the logic" used in automated decision-making. California Consumer Privacy Act of 2018, CAL. STAT. tit. 1.81.5, § 1798.185(a)(16).

indicates that whether explanations improve human decision-making in relation to AI is situationally dependent. First, even when people are given the formulas used by algorithms, they still underperform compared to the algorithm.[324] Second, in a recent series of experiments involving AI recommendations, researchers found that AI systems outperformed human recommenders even in a domain where people have well-developed tastes: predicting what people will find funny.[325] The researchers further found that when these recommender systems outperform friends, family members, and significant others, people still do not trust them.[326] However, the experiments found that trust, and algorithmic preference, did improve when certain explanations were offered about the automated recommendation process.[327] In the explanation condition, subjects were told to

> [t]hink of the algorithm as a tool that can poll thousands of people and ask them how much they like different jokes. This way, the algorithm can learn which jokes are the most popular overall, and which jokes appeal to people with a certain sense of humor. Using the database ratings, the algorithm will search for new jokes that are similar to the ones you liked, and dissimilar to the ones you did not like.[328]

Other studies are showing that, counterintuitively, the more an algorithm is transparent and attempts to be explainable, the more it reduces people's ability to detect and correct model errors, perhaps because of information overload, and does not appear to increase its acceptance.[329] Such lessons about how to build trust through the right framing of explanations could prove an invaluable policy tool.

---

324. Marta Serra-Garcia & Uri Gneezy, *Improving Human Deception Detection Using Algorithmic Feedback* 20–21 (CESifo, Working Paper No. 10518, 2023).

325. Michael Yeomans et al., *Making Sense of Recommendations*, 32 J. BEHAV. DECISION MAKING 403, 403–04 (2019), https://doi.org/10.1002/bdm.2118 [https://perma.cc/X9XF-9JR8].

326. *Id.*

327. *Id.* at 411.

328. *Id.*

329. *See* Forough Poursabzi-Sangdeh et al., *Manipulating and Measuring Model Interpretability*, CHI CONF. ON HUM. FACTORS IN COMPUTING SYS., 2021, at 1, https://dl.acm.org/doi/10.1145/3411764.3445315 [https://perma.cc/G58V-M4TE]. *But see* Daniel Ben David et al., *Explainable AI and Adoption of Financial Algorithmic Advisors: An Experimental Study*, 2021 AAAI/ACM CONF. ON AI, ETHICS & SOC'Y § 4 (2021), https://doi.org/10.48550/arXiv.2101.02555 [https://perma.cc/9EUC-8JPV] (observing that accuracy-based explanations of a model in its initial phase led to higher adoption rates by participants); John Zerilli et al., *How Transparency Modulates Trust in Artificial Intelligence*, 3 PATTERNS 1 (2022), https://doi.org/10.1016/j.patter.2022.100455 [https://perma.cc/EU36-54BQ] (finding that transparency in AI, along with dynamic task allocation, communication of confidence metrics, and other similar strategies, are crucial in promoting trust in AI).

Turning again to the sphere of decisions, some research indicates that people prefer human decision-making to algorithms when moral or subjective questions are at issue. One experiment on moral decision-making showed that people prefer humans, who have discretion, to algorithms, which applied particular human-created fairness principles more consistently than the humans.[330] Other studies show that increasing a task's perceived objectivity increases trust in use of algorithms for that task.[331] Studies also show that people prefer human decision-making in inherently uncertain domains (e.g., medicine and investing).[332] Much like how people want a subjective decision-maker to make subjective decisions, people want an unpredictable decision-maker to make unpredictable decisions, and humans are (mostly rightly) perceived as more subjective and more unpredictable than algorithms.[333] On AI trust in relation to equality, a study of public perception of automated decision-making for bail in California showed that most people believed that an algorithmic decision-maker would increase rather than decrease racial and socioeconomic disparities, and, thus, were unsupportive of automated decision-making.[334] This may well be a predictable finding given the imbalanced coverage of AI-as-Wrongs and algorithmic bias compared to coverage of AI-for-Good and the potential of algorithms to mitigate bias.

On the right to a human-in-the-loop, there is also evidence that despite hopes for effective human-AI decision-making collaboration, humans and AI deciding together often underperform the AI.[335] And yet, one

330. Johanna Jauernig et al., *People Prefer Moral Discretion to Algorithms: Algorithm Aversion Beyond Intransparency*, 35 PHIL. & TECH. 11–12 (2022).

331. Noah Castelo et al., *Task-Dependent Algorithm Aversion*, 56 J. MKTG. RSCH. 809, 823 (2019), https://doi.org/10.1177/0022243719851788 [https://perma.cc/TN4M-XMGW]; *see also* Evan Weingarten et al., *Human Experts Outperform Technology in Creative Markets*, 6 SHE JI: J. DESIGN, ECON., & INNOVATION, 301, 314–16 (2020), https://doi.org/10.1016/j.sheji.2020.07.004 [https://perma.cc/APV8-P3QV] (finding that given current tools, on average, at the same price levels, humans produce more insightful logos, but managers enjoy the process of creating them more with the AI).

332. Dietvorst & Bharti, *supra* note 316, at 1309–11.

333. *Id.* At the same time, in a series of studies by management professor Jennifer Logg and her collaborators, people showed algorithmic appreciation when making numeric estimates about a visual stimulus as well as forecasts about the popularity of songs and romantic matches. Logg et al., *supra* note 300, at 93–95.

334. Nicholas Scurich & Daniel A. Krauss, *Public's Views of Risk Assessment Algorithms and Pretrial Decision Making*, 26 PSYCH., PUB. POL'Y & L. 1, 5 (2020); *see also* Theo Araujo et al., *In AI We Trust? Perceptions About Automated Decision-Making by Artificial Intelligence*, 35 A.I. & SOC'Y 611 (2020) (describing perceptions and realities about the risks, trustworthiness, and fairness of AI reveal gaps).

335. *See, e.g.*, Ben Green & Yiling Chen, *The Principles and Limits of Algorithm-in-the-Loop Decision Making*, 3 PROC. ACM ON HUM.-COMPUT. INTERACTION 4 (2019),

thing that has helped bridge the gap between algorithm aversion and appreciation is control—people who had even a slight amount of control over an algorithm's forecast were more likely to trust the algorithm.[336] In one study, giving participants the freedom to slightly modify the algorithm made them feel more satisfied with the forecasting process, more likely to believe that the algorithm was superior, and more likely to choose to use an algorithm to make subsequent forecasts.[337] Similarly, participants in a different study were more likely to follow algorithmic advice once their forecasts were integrated into the algorithm.[338]

Interestingly, it seems that when the stakes are highest, people are quite willing to trust AI. Take aviation again: the entire commercial flight industry is based on the gold standard that when weather conditions are particularly bad, pilots rely on autopilot.[339] Israel's Iron Dome is almost fully automated, protecting the entire country against missile attacks.[340]

Similarly, people's trust in AI is evident in medicine. A 2021 study about Type 1 diabetes self-management found that participants preferred algorithmic decision-making to human decision-making.[341] The study identified several factors that contribute to algorithm trust, including previous algorithm use and the need for precision.[342] Previous algorithm use by a patient strongly predicts future algorithm use, suggesting a learning process of AI trust, and, generally, as the need for precision increases, algorithm use increases as well.[343] Examining the "bolus calculator" usage behavior in over 306,000 bolus insulin decisions by diabetics, the study also finds that algorithm use declines from morning

https://doi.org/10.1145/3359152 [https://perma.cc/S2H8-W96P]; Sarah Tan et al., *Investigating Human + Machine Complementarity for Recidivism Predictions* (Dec. 3, 2018) (unpublished manuscript), https://doi.org/10.48550/arXiv.1808.09123 [https://perma.cc/BX5T-3MUT].

336. *See* Berkeley J. Dietvorst et al., *Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them*, 64 MGMT. SCI. 1155, 1159, 1161 (2018).

337. *Id.* at 1156, 1165.

338. *See generally* Kohei Kawaguchi, *When Will Workers Follow an Algorithm? A Field Experiment with a Retail Business*, 67 MGMT. SCI. 1670 (2021).

339. *See* Jack Nicas & Zach Wichter, *A Worry for Some Pilots: Their Hands-On Flying Skills Are Lacking*, N.Y. TIMES (Mar. 14, 2019), https://www.nytimes.com/2019/03/14/business/automated-planes.html [https://perma.cc/X4LR-T3YB].

340. Jen Kirby, *Israel's Iron Dome, Explained by an Expert*, VOX (May 14, 2021, 3:40 PM), https://www.vox.com/22435973/israel-iron-dome-explained [https://perma.cc/LSK8-YJ6H].

341. Wilson Lin et al., *What Drives Algorithm Use? An Empirical Analysis of Algorithm Use Determinants in Diabetes Self-Management* 16 (Oct. 13, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3891832 [https://perma.cc/827V-7C5J].

342. *Id.* at 4–5.

343. *Id.* at 24.

to evening, suggesting perhaps that when humans are more tired and cognitively depleted, they are more prone to irrational AI distrust.[344]

In sharp contrast to these examples of high-stakes spheres—national security, health care management, and travel safety—the contemporary policy impulse to insert a human-into-the-loop is also higher when stakes are high. The EU Draft AI Act differentiates between high- and low-risk AI systems, providing that "human oversight shall aim at preventing or [minimizing] the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse."[345] The draft regulation requires high-risk AI systems to be "designed and developed in such a way, including with appropriate human-machine interface tools, that they can be effectively overseen by natural persons."[346] It also bans certain uses of AI that create "unacceptable risk," although it does not specify what those risks are.[347] It then further divides the world of AI risk between high and low, subjecting high-risk AI systems to elaborate risk regulation, all funneled into the abstract requirement of lowering risks to an "acceptable level," although there are no principles in the Act on what are acceptable levels of risks or a requirement to compare such risks with the risks emanating from the status quo of human processing.[348]

It is entirely possible that at least some of the calls for human-in-the-loop rights are agnostic to tangible costs or benefits of adding a human decision-maker. Rather, they may stem from a principle about the inherent value of interacting with a human. Some may believe that people have the right to be in the driver's seat, literally or figuratively, and that no robot should take on a task that humans have been doing for a long time.[349] Similarly, privacy protections are sometimes justified by scholars on grounds that do not consider tangible consequences but refer to dignitarian harms—the inherent immorality of being subjected to surveillance or data extraction. This is part of why biometrics has been

344. *Id.* at 9. The researchers in this study do not contemplate this aspect, but on cognitive depletion and fatigue, see Anastasiya Pocheptsova et al., *Deciding Without Resources: Resource Depletion and Choice in Text*, 46 J. MKTG. RSCH. 344, 348–49 (2009).

345. EU Draft AI Act, *supra* note 15, at 51.

346. *Id.*

347. *Id.* at 12–13.

348. *Id.*

349. Relatedly, human decision-makers, such as referees in athletic matches or judges in the courtroom, serve not only in handing down tangible decisions but also play a performative role. *See* Michael J. Madison, *Fair Play: Notes on the Algorithmic Soccer Referee*, 23 VAND. J. ENT. & TECH. L. 341, 372–74 (2021), https://scholarship.law.vanderbilt.edu/jetlaw/vol23/iss2/4 [https://perma.cc/WF9M-ZHDP] (relating to the algorithmic trust question discussed in Section III.C).

especially controversial.[350] There may be moral reasons to want to take these intangible values into account. This Article does not deny or defend such a possibility. Rather, this Article urges policy debates to be clear about when such reasons are being invoked, what work they are doing in the analysis, and what the costs are that we, as a society, are willing to pay to prioritize them over tangible harms—for example, inaccuracy, inequity, or resource depletion. Even if there are such reasons to privilege human action and privacy, the questions about how to foster rational trust in AI needs to be studied more and not just the effectiveness and accuracy of the AI models themselves.[351]

Related to these dilemmas, behavioral research can help unpack the driving forces behind the right to a human-in-the-loop. For example, can the benefits of interacting with a human be achieved by making AI appear human(oid)? In one experiment, giving an autonomous vehicle more human-like features led to more trust in it.[352] Those who drove an autonomous vehicle that was named, gendered, and voiced reported trusting their vehicle more, were more relaxed in an accident and blamed their vehicle and related entities less for an accident caused by another driver.[353] The authors suggest that blurring the line between humans and machines could increase users' willingness to trust technology in place of humans.[354] Policy studies of human interactions with the physical design of robots must complement the behavioral research of human trust in digital unembodied AI.[355]

CONCLUSION

Automation has the potential to increase accuracy, inclusion, fairness, access, and efficiency in areas ranging from health to education, from climate change to poverty. Including an AI-for-Good policy agenda on an equal footing with the prevalent AI-as-Risk regulatory framework

350. *See* Lobel, *Biopolitical Opportunities*, *supra* note 291, 192–93.

351. Lessons from past tech revolutions may be valuable. *See, e.g.*, Adrienne LaFrance, *When People Feared Computers*, ATLANTIC (Mar. 30, 2015), https://www.theatlantic.com/technology/archive/2015/03/when-people-feared-computers/388919/ [https://perma.cc/PEV2-5D3J] ("In the early 1980s, the age of the personal computer had arrived and 'computerphobia' was suddenly everywhere. Sufferers experienced 'a range of resistances, fears, anxieties, and hostilities,' according to the 1996 book *Women and Computers*.").

352. *See* Adam Waytz et al., *The Mind in the Machine: Anthropomorphism Increases Trust in an Autonomous Vehicle*, 52 J. EXPERIMENTAL SOC. PSYCH 113, 116 (2014).

353. *Id.*

354. *Id.* In my book *The Equality Machine*, I further explore the value, costs and benefits, and risks and normative dilemmas of anthropomorphizing AI, especially regarding gendering chatbots and humanoids. LOBEL, *supra* note 29, at chs. 9–10.

355. *See, e.g.*, Sonja K. Ötting et al., *Let's Work Together: A Meta-Analysis on Robot Design Features That Enable Successful Human–Robot Interaction at Work*, 64 HUM. FACTORS 1027, 1028 (2022).

brings to the forefront the ways policy reforms have been narrow and limited. Moreover, it illuminates the many roles policy can play in supporting and directing positive change. Behavioral research has long illuminated biases and cognitive failures we humans are susceptible to due to our black box algorithms called brains.[356] Adopting a comparative lens that considers both human fallibility and the flaws and advantages of new digital technology is a superior way for policymakers to critically assess and positively support the rapid developments we will face as a society in the near future.

356. Yuval Feldman & Orly Lobel, *Behavioral Tradeoffs: Beyond the Land of Nudges Spans the World of Law and Psychology*, *in* NUDGE AND THE LAW: A EUROPEAN PERSPECTIVE 125 (A. Alemanno & A. Sibony eds., 2015); Lobel & Amir, *Liberalism and Lifestyle*, *supra* note 265, at 20–21.