REGULATING THE RISKS OF AI

MARGOT E. KAMINSKI[*]

ABSTRACT

Companies and governments now use Artificial Intelligence ("AI") in a wide range of settings. But using AI leads to well-known risks that arguably present challenges for a traditional liability model. It is thus unsurprising that lawmakers in both the United States and the European Union ("EU") have turned to the tools of risk regulation in governing AI systems.

This Article describes the growing convergence around risk regulation in AI governance. It then addresses the question: what does it mean to use risk regulation to govern AI systems? The primary contribution of this Article is to offer an analytic framework for understanding the use of risk regulation as AI governance. It aims to surface the shortcomings of risk regulation as a legal approach, and to enable readers to identify which type of risk regulation is at play in a given law. The theoretical contribution of this Article is to encourage researchers to think about what is gained and what is lost by choosing a particular legal tool for constructing the meaning of AI systems in the law.

Whatever the value of using risk regulation, constructing AI harms as risks is a choice with consequences. Risk regulation comes with its own policy baggage: a set of tools and troubles that have emerged in other fields. Risk regulation tends to try to fix problems with the technology so it may be used, rather than contemplating that it might sometimes not be appropriate to use it at all. Risk regulation works best on quantifiable problems and struggles with hard-to-quantify harms. It can cloak what are really policy decisions as technical decisions. Risk regulation typically is not structured to make injured people whole. And the version of risk regulation typically deployed to govern AI systems lacks the feedback loops of tort liability. Thus the choice to use risk regulation in the first place channels the law towards a particular approach to AI governance that makes implicit tradeoffs and carries predictable shortcomings.

The second, more granular observation this Article makes is that not all risk regulation is the same. That is, once regulators choose to deploy risk regulation, there are still significant variations in what type of risk regulation they might use. Risk regulation is a legal transplant with multiple possible origins. This

[*] Associate Professor of Law, Colorado Law School. Director of the Privacy Initiative, Silicon Flatirons Center. Affiliated Faculty, Information Society Project at Yale Law School. Thanks in particular to Eric Hilgendorf, Sharon Jacobs, Meg Leta Jones, Matt Sag, Daniel Solove, Tal Zarsky, and to participants in the Colorado Law School Juniors Workshop and We Robot for either workshopping drafts or pointing me to relevant literature. Mistakes are my own.

Article identifies at least four models for AI risk regulation that meaningfully diverge in how they address accountability.

CONTENTS

INTRODUCTION

Companies and governments are increasingly using Artificial Intelligence ("AI") in a wide range of applications, including in "critical decisions about Americans' health, finances, housing, educational opportunities and more."[1] AI systems are touted as more efficient, less expensive, and potentially more accurate and less biased than human decision makers.[2] But using AI has well-known risks.[3] Software can crash.[4] Outcomes can be unpredictable, even irrational.[5] AI systems can be overfit or underfit to their training data.[6] An AI system that works well in one situation may fail egregiously in another.[7]

[1] See Press Release, Ron Wyden, U.S. Senator from Oregon, Wyden, Booker and Clarke Introduce Algorithmic Accountability Act of 2022 to Require New Transparency and Accountability for Automated Decision Systems, Feb. 3, 2022, https://www.wyden.senate.gov/news/press-releases/wyden-booker-and-clarke-introduce-algorithmic-accountability-act-of-2022-to-require-new-transparency-and-accountability-for-automated-decision-systems.

[2] *See, e.g.*, Alex P. Miller, *Want Less-Biased Decisions? Use Algorithms.*, HARV. BUS. REV. (July 26, 2018), https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms [https://perma.cc/9QDS-QUFX] ("Algorithms are less biased and more accurate than the humans they are replacing."). *But see* Stephanie Bornstein, *Antidiscriminatory Algorithms*, 70 ALA. L. REV. 519, 521 (2018) (explaining that to learn what to value in its algorithm, AI must be programed by humans who unknowingly insert biases).

[3] There is disagreement over how to define "Artificial Intelligence." *See, e.g.*, Harry Surden, *Artificial Intelligence and Law: An Overview*, 35 GA. ST. U. L. REV. 1305, 1310 (2019) (discussing how AI is interdisciplinary, rather than just computer science); *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*, annexes at 1, COM (2021) 206 final (Apr. 21, 2021) [hereinafter *Draft EU AI Act*] (listing AI techniques).

[4] *See, e.g.*, Bryan H. Choi, *Crashworthy Code*, 94 WASH. L. REV. 39, 64 (2019) (stating software errors occur because it is computationally impossible to find all bugs due to magnitude of code). Or, problems can arise elsewhere, such as in how an AI senses the outside world. *See, e.g.*, Jason Torchinsky, *Watch Teslas and Audis and Other Cars Smack into Fake Kids in Demos of Automatic Emergency Braking Systems*, JALOPNIK (Jan. 6, 2022), https://jalopnik.com/watch-teslas-and-audis-and-other-cars-smack-into-fake-k-1848316218 [https://perma.cc/XM5K-AA35] (discussing findings of lidar company's testing of self-driving cars).

[5] *See generally* JANELLE SHANE, YOU LOOK LIKE A THING AND I LOVE YOU: HOW ARTIFICIAL INTELLIGENCE WORKS AND WHY IT'S MAKING THE WORLD A WEIRDER PLACE (2019) (demonstrating erratic way AI machines learn and produce outputs).

[6] *See, e.g.*, NIST, AI RISK MANAGEMENT FRAMEWORK (AI RMF 1.0) 13 (2023) [hereinafter NIST, AI RMF].

[7] *See, e.g.*, Sasha Harrison, *How To Tame the Long Tail in Machine Learning*, SCALE (June 29, 2021), https://scale.com/blog/taming-long-tail [https://perma.cc/5ZF4-VS4A] (showing difficulty in getting AI "to perform well on examples not adequately represented in the training dataset"); Andrew D. Selbst, danah boyd, Sorelle A. Friedler, Suresh Venkatasubramanian & Janet Vertesi, *Fairness and Abstraction in Sociotechnical Systems*, FAT* '19, Jan. 2019, at 59, 59 (explaining when technology is added to our social system it has both expected and unexpected effects).

Complex human-machine systems have well-known points of weakness and are vulnerable to failure cascades.[8] And using AI systems may actively discriminate against members of protected classes or more subtly replicate and perpetuate existing societal biases.[9]

It is thus at first glance unsurprising that lawmakers in both the United States and the European Union ("EU") have turned to the tools of risk regulation to govern AI systems. In the United States, legislators at both the state and federal level have proposed requiring risk assessments and risk mitigation for certain uses of AI systems.[10] The U.S. Department of Commerce's National Institute of Standards and Technology ("NIST") has developed an Artificial Intelligence Risk Management Framework ("AI RMF").[11] The Draft EU AI Act is built on the existing scaffolding of product safety risk regulation.[12] The EU's General Data Protection Regulation ("GDPR") deploys risk regulation alongside its much-discussed individual rights.[13] This Article is the first to examine and compare this set of laws and proposed laws and place them in the context of broader discussions of risk regulation, including in other regulated fields.

This Article identifies that AI governance has been converging around a particular model of risk regulation. It then asks what this convergence means. Risk regulation is neither inevitable nor a neutral tool. Framing the potential harms of AI systems as *risks* and the solutions as *risk regulation* are value-laden choices. The concept of risk itself brings with it a distinct perspective and normative leanings. There are many possible ways to legally address danger and harm, from precautionary bans to class action lawsuits. The lighter-touch risk regulation currently being deployed in AI governance carries with it a particular

[8] Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 VAND. L. REV. 429, 438 (2023) (noting inadequate training, interface issues, and bungled handoffs as weaknesses in human-led systems).

[9] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 673-74 (2016) (exploring how institutional discrimination can be unknowingly programmed into AI); Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2221-24 (2019) (exploring algorithmic bias in criminal justice).

[10] Andrew D. Selbst, *An Institutional View of Algorithmic Impact Assessments*, 35 HARV. J.L. & TECH. 117, 117-21 (2021) [hereinafter Selbst, *Algorithmic Impact Assessments*]; S.B. 5116, 67th Leg., Reg. Sess. (Wash. 2021); Algorithmic Accountability Act of 2022, H.R. 6580, 117th Cong. (2022).

[11] NIST, AI RMF, *supra* note 6, at 2.

[12] *Draft EU AI Act*, *supra* note 3, at 13; *see also* Michael Veale & Frederik Zuiderveen Borgesius, *Demystifying the Draft EU Artificial Intelligence Act*, 22 COMPUT. L. REV. INT'L 97, 97 (2021).

[13] Commission Regulation 2016/679, 2016 O.J. (L 119) 1 [hereinafter GDPR]; *see also* Claudia Quelle, *The 'Risk Revolution' in EU Data Protection Law: We Can't Have Our Cake and Eat It, Too*, *in* 10 DATA PROTECTION AND PRIVACY: THE AGE OF INTELLIGENT MACHINES 33, 33 (Ronald Leenes, Rosamunde van Brakel, Serge Gutwirth & Paul De Hert eds., 2017) [hereinafter Quelle, *The 'Risk Revolution'*] (discussing potential conflict between rights-based and risk-based approaches).

way of framing problems and harms and a limited set of regulatory tools for addressing them.

Thus, this Article asks: what does it mean to legally construct the harms of AI systems through risk regulation?[14] By framing the regulation of AI systems as risk regulation, policymakers are, knowingly or not, taking a normative stance on AI. First, risk regulation typically assumes a technology will be adopted despite its harms. Second, while aspects of risk regulation may be effective at certain kinds of harm mitigation, risk regulation as a legal interface elides, or renders invisible, both certain kinds of harms (typically, those that are less readily quantifiable) and certain individuals and populations (typically, marginalized individuals and populations) harmed by AI.

The primary contribution of this Article is thus to offer an analytic framework for understanding the use of risk regulation as AI governance. It aims to surface the shortcomings of risk regulation as a legal approach, and to enable readers to identify the type of risk regulation at play in a given law or regulation. At this level, this Article contributes to the immediate policy discussion. It situates the developing risk regulation of AI systems against risk regulation in other fields, such as environmental law, health law, and safety regulation. It points out known shortcomings of risk regulation, and the ways in which other areas of law address or mitigate them.

The second, more theoretical, contribution of this Article is to an ongoing debate in the field of law and technology over whether technology drives the law or the law constructs technology.[15] This Article takes the position that the legal system itself plays a significant role in its own encounters with sociotechnical change.[16] That is, the law doesn't passively react to

[14] This Article uses the methodology of "legal construction." *See* MEG LETA JONES, THE CHARACTER OF CONSENT: THE HISTORY OF COOKIES AND FUTURE OF TECHNOLOGY POLICY (MIT Press, forthcoming 2023); Meg Leta Jones, *Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw*, 18 U. ILL. J.L. TECH. & POL'Y 249, 281 (2018) [hereinafter Jones, *Tech Exceptionalism*] ("I use the term legal construction of technology . . . to tie the 'social construction of technology' in [science and technology studies] to the field of law and technology."); Margot E. Kaminski, *Authorship, Disrupted: AI Authors in Copyright and First Amendment Law*, 51 U.C. DAVIS L. REV. 589, 593 (2017) [hereinafter Kaminski, *Authorship Disrupted*] (defining legal construction as "understanding how the law makes meaning of technologies").

[15] *See* Ryan Calo, *Robotics and the Lessons of Cyberlaw*, 103 CALIF. L. REV. 513, 515 (2015) (exploring issues that may arise if technology changes enough to render current legal system inapplicable); Rebecca Crootof & BJ Ard, *Structuring Techlaw*, 34 HARV. J.L. & TECH. 347, 349 (2021) ("The fundamental challenge of techlaw is not how to best regulate novel technologies, but rather how to best address familiar forms of legal uncertainty in new sociolegal contexts."); Jones, *Tech Exceptionalism*, *supra* note 14, at 250-55 (explaining law has often struggled to process new technology changes); Kaminski, *Authorship Disrupted*, *supra* note 14, at 589-90 (describing how technology is usually seen as outside of law).

[16] Jones, *Tech Exceptionalism*, *supra* note 14, at 253 ("Theories of technological change not only shape the way in which we see social, policy, and legal problems but also the way in which we approach describing, analyzing, and solving such problems.").

sociotechnical change; it constructs it.[17] Deploying a particular legal interface such as risk regulation changes how the social practices of AI are read into the legal system. The developing law of AI plays an active role in legally constructing the meaning of AI systems and their accompanying harms.

This Article proceeds as follows: Part I makes the descriptive contribution that regulators have been converging around risk regulation in AI governance. Past scholarship has focused on the use of specific tools such as audits and impact assessments in AI governance.[18] Recent scholarship more explicitly calls for the use of risk regulation to govern AI.[19] But as of yet, researchers have not examined proposed and enacted laws that use risk regulation to govern AI systems. Part I begins by explaining what AI systems are, and what harms they

[17] Kaminski, *Authorship Disrupted*, *supra* note 14, at 590-91 ("The law, in constructing—that is, building the meaning of—new technological developments and their social uses, takes a central part in its own disruption.").

[18] *See, e.g.*, Selbst, *Algorithmic Impact Assessments*, *supra* note 10, at 122-27; Margot E. Kaminski & Gianclaudio Malgieri, *Algorithmic Impact Assessments Under the GDPR: Producing Multi-Layered Explanations*, 11 INT'L DATA PRIV. L. 125, 125 (2021) (discussing data protection impact assessments ("DPIAs") under the GDPR). For other scholarship discussing DPIAs, see Reuben Binns, *Data Protection Impact Assessments: A Meta-Regulatory Approach*, 7 INT'L DATA PRIV. L. 22, 29-30 (2017); Bryan Casey, Ashkon Farhangi & Roland Vogl, *Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise*, 34 BERKELEY TECH. L.J. 143, 171-77 (2019); Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking for*, 16 DUKE L. & TECH. REV. 18, 77-80 (2017). Civil society groups have also written about impact assessments. *See, e.g.*, EMANUEL MOSS, ELIZABETH ANNE WATKINS, RANJIT SINGH, MADELEINE CLARE ELISH & JACOB METCALF, DATA & SOC'Y, ASSEMBLING ACCOUNTABILITY: ALGORITHMIC IMPACT ASSESSMENT FOR THE PUBLIC INTEREST 7 (2021); DILLON REISMAN, JASON SCHULTZ, KATE CRAWFORD & MEREDITH WHITTAKER, AI NOW, ALGORITHMIC IMPACT ASSESSMENTS: A PRACTICAL FRAMEWORK FOR PUBLIC AGENCY ACCOUNTABILITY 4-6 (2018). For the notion that impact assessments should reflect not only AI developers' understanding of risks but also the public's, see Jacob Metcalf, Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh & Madeleine Clare Elish, *Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts*, FAcct '21, Mar. 2021, at 735, 735. For a discussion of risk regulation under the GDPR, see Quelle, *The 'Risk Revolution,' supra* note 13, at 35; Claudia Quelle, *Enhancing Compliance Under the General Data Protection Regulation: The Risky Upshot of the Accountability- and Risk-Based Approach*, 9 EURO. J. RISK REGUL. 502, 502 (2018); RAPHAËL GELLERT, THE RISK-BASED APPROACH TO DATA PROTECTION (2020); Raphaël Gellert, *Understanding the Notion of Risk in the General Data Protection Regulation,* 34 COMPUT. L. & SEC. REV. 279, 279 (2018). For a discussion of risk regulation and autonomous vehicles, see generally Matthew T. Wansley, *Regulation of Emerging Risks*, 69 VAND. L. REV. 401 (2016).

[19] *See, e.g.*, Michael Guihot, Anne F. Matthew & Nicolas P. Suzor, *Nudging Robots: Innovative Solutions To Regulate Artificial Intelligence*, 20 VAND. J. ENT. & TECH. L. 385, 445 (2017); Gary E. Marchant & Yvonne A. Stevens, *Resilience: A New Tool in the Risk Governance Toolbox for Emerging Technologies*, 51 U.C. DAVIS L. REV. 233, 236 (2017); Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L & TECH. 353, 356 (2016); Alicia Solow-Niedermann, *Administering Artificial Intelligence*, 93 S. CAL. L. REV. 633, 691-93 (2020).

can cause. It identifies arguments for the turn to risk regulation. Finally, Part I identifies shared features across laws and policy proposals, including transatlantic examples.

Part II then asks the question: what does it mean to legally construct AI systems through risk regulation? It points to several predictable consequences of the choice to go the risk regulation route. Part II identifies what I call the "policy baggage" of risk regulation: a set of practices and problems that predictably emerge when risk regulation is deployed. Risk regulation is what Jessica Eaglin has termed "techno-correctionist" by nature, largely trying to "fix" the technology so that it might be used, rather than taking as a starting point that sometimes it might be better not to deploy the technology at all.[20] Risk regulation works best on problems that are easy to quantify, struggling with unquantifiable harms. It can cloak what are really policy decisions as technical decisions. It may fail to provide civil recourse, including compensation. And the lighter-touch version of risk regulation increasingly being deployed to govern AI systems lacks the responsiveness of tort liability, with efforts to make risk regulation more responsive subject to capture by regulated industries.

Part III then observes that choosing to use risk regulation is not the end of the story. Policymakers also choose what kind of risk regulation to apply. There are, in fact, multiple models of risk regulation, as risk regulation is a legal transplant with multiple different origins.[21] Part III identifies at least four different versions of risk regulation: the heavily quantitative version of risk assessment that emerged in U.S. administrative law in the 1960s-1980s; risk regulation as democratic oversight (i.e., the model used in NEPA); a "UK Version" of risk regulation, which on a macro level allocates regulatory capacity and enforcement based on the risks posed; and enterprise risk management, exemplified by the standards offered by NIST. These models diverge around how soft or hard the law is, and around how to structure accountability. This Part aims to enable researchers and policymakers to better understand the consequences of deploying a particular form of risk regulation to regulate AI.

I. THE CONVERGENCE AROUND AI RISK REGULATION

Lawmakers around the world have been converging around a particular model of AI governance: risk regulation. This Part identifies this trend. It begins by explaining what AI systems are and what harms concern regulators. It then identifies the arguments for using largely ex ante risk regulation, as opposed to focusing on ex post liability. Finally, this Part closes by identifying what AI risk

[20] Jessica M. Eaglin, *When Critical Race Theory Enters the Law & Technology Frame*, 26 MICH. J. RACE & L. (SPECIAL ISSUE) 151, 155-57 (2021).

[21] This discussion brings concepts from the legal transplants literature into conversation with the risk-regulation literature. Vanessa Casado Pérez & Yael R. Lifshitz, *Natural Transplants*, 97 N.Y.U. L. REV. 933, 935 (2022) (explaining legal transplant is "a transfer of a legal regime or rule from one jurisdiction to another").

regulation looks like, pointing to common features that have emerged across different laws in different jurisdictions.

A. *The Risks of AI*

There is an ongoing debate over how to define AI.[22] Proposed laws typically characterize AI systems as software systems with autonomy: computer programs that can produce outputs such as content or predictions with minimal human involvement.[23] There is debate over whether to aim regulation at particular kinds of software programs, such as machine learning programs, or to more broadly future-proof laws by covering types of software some might struggle to identify as AI.[24] In general, AI systems work by scanning huge data sets and drawing predictions and conclusions from them that are then applied to create predictions or outputs.[25] Even within the category of machine learning algorithms, AI systems can be built and trained in a variety of meaningfully different ways.[26]

[22] *See, e.g.*, Bryan Casey & Mark A. Lemley, *You Might Be a Robot*, 105 CORNELL L. REV. 287, 287 ("No one has been able to offer a decent definition of robots and AI—not even experts.") (2020).

[23] *Draft EU AI Act*, *supra* note 3, at 18 ("The definition should be based on the key functional characteristics of the software, in particular the ability, for a given set of human-defined objectives, to generate outputs such as content, predictions, recommendations, or decisions which influence the environment with which the system interacts, be it in a physical or digital dimension."); NIST, AI RMF, *supra* note 6, at 1 (Defining an "AI system as an engineered or machine-based system that can, for a given set of human-defined objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments.").

[24] For example, the 2021 draft version of the proposed AI Act, in an attempt to future proof the law against new technologies and new uses, defined AI broadly as software developed with machine learning, logic- and knowledge-based, and/or statistical approaches. *Draft EU AI Act*, *supra* note 3, annexes at 1. More recently, the EU Parliament proposed amending the AI Act's definition to more narrowly focus on autonomous systems, and largely match the definition of AI used by the Organization for Economic Cooperation and Development: a "system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate output such as predictions, recommendations, or decisions influencing physical or virtual environments." Luca Bertuzzi, *EU Lawmakers Set To Settle on OECD Definition for Artificial Intelligence*, EURACTIV (Mar. 9, 2023), https://www.euractiv.com/section/artificial-intelligence/news/eu-lawmakers-set-to-settle-on-oecd-definition-for-artificial-intelligence/ [https://perma.cc/N2J3-4WEM]; *see also* Luca Bertuzzi, *AI Act: All the Open Political Questions in the European Parliament*, EURACTIV (Feb. 15, 2023), https://www.euractiv.com/section/artificial-intelligence/news/ai-act-all-the-open-political-questions-in-the-european-parliament/ [https://perma.cc/FJ7H-J8S2] (explaining EU lawmakers' earlier proposed using definition from NIST).

[25] David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 671 (2017).

[26] *Id.* at 658.

Importantly, for all the hype, AI systems can and do fail. I and others have catalogued the harms and benefits of AI systems at length elsewhere.[27] I have claimed that concerns motivating the regulation of AI can be dignitary, justificatory, or instrumental in nature—or often, some mix of the three.[28] Largely, AI risk regulation is motivated by instrumental concerns: the goal of making AI systems avoid harms and work better. An instrumental approach to AI regulation focuses on ensuring that AI systems are more accurate and avoid systemic bias.

Readers familiar with the risks and harms of AI systems should feel free to skim or skip this Section and move on to Section I.B below, which outlines the arguments for risk regulation. This Section provides three case studies to more concretely illustrate the harms of AI systems in a variety of contexts. These case studies demonstrate that AI systems are no longer hypothetical but are in wide use across many sectors. They illustrate the challenges and the appeal of taking an omnibus, technology-specific approach to regulating AI, as the EU has done.[29]

AI harms are very different in different contexts, where they might be already addressed by particular sectoral laws. At the same time, most AI harms can readily be traced to a pattern of similar problems. Problems of incomplete or biased training data ("garbage in, garbage out") and poorly designed human-machine systems (including ignoring known cognitive biases and overreliance on a human in the loop)[30] resonate across sectors.

The first case study addresses self-driving cars; the second, AI and employee recruitment; and the third, AI and public health. In each example, I aim to surface similarities across contexts. I also aim to illustrate the difference between

[27] *See, e.g.*, Margot E. Kaminski, *Binary Governance: Lessons from the GDPR's Approach to Algorithmic Accountability*, 92 S. CAL. L. REV. 1529, 1529, 1533 (2019) [hereinafter Kaminski, *Binary Governance*] (explaining why individual rights and public- and stakeholder-facing accountability are crucial to regulating algorithmic decision making); Crootof et al., *supra* note 8, at 452 (describing several ways in which algorithmic systems can cause harm, including automated weapons killing civilians, autonomous vehicles injuring pedestrians, and medical imaging systems misidentifying tumors); Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671, 1679 (2020) [hereinafter Ajunwa, *Paradox of Automation*] (arguing algorithmic bias in hiring requires legal, not technical, solution); Solow-Niederman, *supra* note 19, at 633 ("[T]he current trajectory of Al development, which is dominated by large private firms, augurs an era of private governance."); Barocas & Selbst, *supra* note 9, at 684 (highlighting "systematic disadvantage that members of protected classes may suffer from being miscounted and, as a result, misrepresented in the evidence base"); Aziz Z. Huq, *A Right to Human Decision*, 106 VA. L. REV. 611, 619 (2020) (proposing that in the future "well-calabrated" AI may be preferrable to human decision making); Cary Coglianese & Alicia Lai, *Algorithm vs. Algorithm*, 71 DUKE L.J. 1281, 1281 (2022) ("Digital algorithms, such as machine learning, can improve governmental performance by facilitating outcomes that are more accurate, timely, and consistent.").

[28] Kaminski, *Binary Governance*, *supra* note 27, at 1534.

[29] *Draft EU AI Act*, *supra* note 3, at 3.

[30] *See generally* Crootof et al., *supra* note 8.

characterizing the harms caused by AI systems as individualized harms and looking at them as systemic risks. This illustrates both the appeal of the turn to risk regulation, and what is potentially lost in that turn.

1. AI and Safety: Self-Driving Cars

This first case study addresses AI systems in driverless cars.[31] The first documented instance of a self-driving car killing a pedestrian occurred in March 2018.[32] An Uber employee named Rafaela Vasquez spent nine months working as the human "safety operator," or backup driver, of a self-driving Volvo SUV. On a clear night in Tempe, Arizona, the car failed to detect a pedestrian. It did not alert Vasquez to any problem until 0.2 seconds before impact, when she took manual control. The car struck and killed 49-year-old Elaine Herzberg.

For context, car accidents involving human drivers kill more than 38,000 people a year.[33] Self-driving cars are widely touted as a way to make driving safer—to avoid human drivers' inattention, frailties, and errors.[34] Self-driving cars also, however, allow companies like Uber to avoid the costs of employing humans. In 2015, Uber joined the race with Google's Waymo and others to prototype, test, and deploy self-driving cars, with the explicit goal of replacing its human workforce. Arizona paved the way with permissive regulation, and by late 2017, Uber had forty test cars in Arizona running eight shifts a day, trying to rack up test miles.

Initially, Uber employed two human safety operators in each car. One person would call out obstacles, while the other would confirm on a laptop whether the AI system had "seen" them. But in the fall of 2017, Uber changed its policies. One driver now operated the car solo and entered feedback via a tablet mounted on the car dashboard.

These solo operators typically got bored without a second person in the car, and increasingly turned to looking at their phones over long shifts while the car was in motion. Despite firing an increasing number of employees who were caught looking at their phones, Uber did not alter the length of employee shifts and had no policy in place of policing for driver phone use. That is, Uber did not try to systemically avoid the known problem of "automation complacency"—

[31] For a discussion of recent efforts to regulate driverless cars, see generally Matthew Wansley, *Regulating Automated Driving*, 73 EMORY L.J. (forthcoming 2024) (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4190688).

[32] Lauren Smiley, *'I'm the Operator': The Aftermath of a Self-Driving Tragedy*, WIRED (Mar. 8, 2022, 6:00 AM) [hereinafter Smiley, *'I'm the Operator'*], https://www.wired.com/story/uber-self-driving-car-fatal-crash/.

[33] *Id.* (noting human error partly responsible for over ninety percent of crashes).

[34] *Id.* ("By taking sleepiness, inattention, drunkenness, and rage out of the equation and replacing them with vigilant, precise technology, self-driving cars promise to make the roads dramatically safer.").

the tendency of human operators to overly trust effective automated systems.[35] The company also made the decision to override Volvo's automated braking system to avoid false alarms, presumably to increase the number of test miles driven.

Video of Vasquez, the human operator the night of the fatal crash, seemed to indicate that she was looking at her phone in the seconds before the car crashed. Investigators later confirmed that she had been streaming a video during the crash, although defense attorneys have argued that she was listening to audio but not viewing the video stream. Defense attorneys have also argued that Vasquez was looking down at her work phone to check Slack, as required by Uber, her employer. Vasquez was charged with negligent homicide with a dangerous instrument. She pled guilty to endangerment in July 2023 and was sentenced to probation.[36]

This case study exemplifies two ways of looking at AI harms. Through an individualized ex post lens, the car accident, arguably immediately attributable to Vasquez's negligence, caused an individualized and concrete harm: it killed Herzberg. Prosecutors charged Vasquez with negligent homicide, and Uber settled with Herzberg's family. Through the lens of risk, however, the case exemplifies systemic failure: risk mitigation gone awry. Uber's safety culture failed. Its technology failed. Its hybrid human-machine system was poorly designed.[37]

Through the risk lens, then, the story is different. It is not solely about one human backup driver failing to intervene in the actions of a failed machine. The risk lens—the one deployed by the National Transportation Safety Board

[35] "Automation complacency" is defined as "where the user may come to rely completely on the automation even if it is faulty or unreliable." Liz Carver & Murray Turoff, *Human-Computer Interaction: The Human and Computer as a Team in Emergency Management Information Systems*, COMMC'NS. ACM, Mar. 2007, at 33, 38; *see also* Raja Parasuraman & Dietrich H. Manzey, *Complacency and Bias in Human Use of Automation: An Attentional Integration*, 52 HUM. FACTORS & ERGONOMIC SOC'Y 381, 397 (2010) (noting automation complacency cannot be overcome with practice or training); Eugénie Avril, Jordan Navarro, Liên Wioland, Benoît Valery, Virginie Govaere, Didier Gourc, Koosha Khademi, Christos Dimopoulos, Elisabeth Dargent, Nathalie Renaudeau & Julien Cegarra, *Automation and Complacency: Insights from a Planning Task in the Transportation Domain*, *in* HCI INTERNATIONAL 2018—POSTERS' EXTENDED ABSTRACTS 437, 438 (Constantine Stephanidis ed., 2018) (identifying automation complacency in transportation increases with reliability of system). Other companies, such as Google, have recognized the problem of automation complacency and changed their testing protocols accordingly. Five days before the fatal 2018 crash, an Uber operations manager concerned about solo human operators and distractions emailed company executives urging them to reinstate the second driver and cut the fleet size. Smiley, *'I'm the Operator*,*' supra* note 32.

[36] Lauren Smiley, *The Legal Saga of Uber's Fatal Self-Driving Car Crash Is Over*, WIRED (July 28, 2023, 6:47 PM) https://www.wired.com/story/ubers-fatal-self-driving-car-crash-saga-over-operator-avoids-prison/.

[37] *Cf.* Crootof et al., *supra* note 8, at 498 ("Well-designed interfaces are critical. A poorly designed or inadequate interface creates opportunities for critical information to be garbled or lost in translation.").

("NTSB"), by Volvo, and even to some extent by Uber itself—looks towards the future, aiming at preventing more harms. It looks at the aggregate (the safety of self-driving cars in general), and at the systemic level (the technical and organizational systems that lead to the crash). It involves quantifiable testing—such as Volvo's testing that showed that had Uber not disabled the automated braking system, the crash would have been avoided seventeen out of twenty times.[38]

The risk framing, too, breaks down the illusion of simple causality. Instead of "distracted human operator causes pedestrian death," the admittedly less exciting risk-framed headline would read: "Uber's poor safety culture, faulty technology, poor attention to the interface between its system and its drivers, and careless treatment of its employees predictably led to a high risk of crashes." That is, the risk framing looks not so much at the individual crash as at the human-machine system: how the technology operated (badly), how it handed off control to the driver (poorly), and how Uber utterly failed to mitigate, or even worsened, the known problem of automation complacency.[39]

In many ways, as I've argued elsewhere, this reframing of AI harm is descriptively correct.[40] However, it also brings a lot of baggage. For example, framing the harms of pedestrian fatalities as risks means we tacitly accept that we are going to adopt self-driving cars, and that there will be human deaths as a result. There might be fewer human deaths than with human drivers, but they will be a different kind of human death, the result of different kinds of flaws in the system.

2. AI and Employee Recruitment

The second case study involves the use of AI in employee recruitment.[41]

Weeding through job applications is costly and time consuming and exactly the kind of thing that companies would prefer to automate. Many large companies already use AI tools in recruitment, including McDonald's, JP

[38] Smiley, *'I'm the Operator,' supra* note 32.

[39] Uber's own investigation of the crash indicated the technology never identified Herzberg as a person. The AI system had been programmed by Uber to override Volvo's own automated braking system, and thus not to brake hard unless the system was sure it could entirely avoid a crash. In the NTSB's final report on the crash, it named Vasquez's inattention as the immediate cause of the crash but called it a "typical effect of automation complacency" and a reflection in general of Uber's "inadequate safety culture." NTSB chair Robert Sumwalt explained that "[t]he collision was the last link of a long chain of actions and decisions made by an organization that unfortunately did not make safety the top priority." *Id.*

[40] Crootof et al., *supra* note 8, at 487-88 (arguing for systemic approach to regulation).

[41] For more scholarship discussing problems with AI and employment, see, for example, Ifeoma Ajunwa, *Automated Video Interviewing as the New Phrenology*, 36 BERKELEY TECH. L.J. 1173, 1224-25 (2021); Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189, 202 (2017).

Morgan, Kraft Heinz, Deloitte, and LinkedIn.[42] As many as ninety percent of Fortune 500 companies use automation of some kind to screen or rank job candidates.[43]

The use of these AI recruitment systems comes with harms. In 2014, Amazon infamously built a recruitment program to automate its search for talent.[44] It trained the program on resumes submitted over a ten-year period. In 2015, the team noticed that the AI system was biased against women. It downgraded resumes that included the word "women's" and penalized graduates of all-women's colleges.[45] The AI system had replicated the bias in the data set, in which most successful job applicants over the past ten years had been male. This is a classic illustration of the problem of "garbage in, garbage out," in which the quality of machine learning models "is only as good as the quality of [their training] data."[46] Amazon made the decision not to use the AI.

Yet the use of AI systems in recruitment is only growing.[47] Companies that adopt these algorithms often argue that they will be more objective and less biased than humans.[48] But hiring algorithms have been found, like humans, to penalize applicants for having a Black-sounding name.[49] Some algorithms alter

[42] Milena Mikael-Debass, *Companies Are Using AI To Stop Bias in Hiring. They Could Also Make Discrimination Worse.*, VICE (Dec. 28, 2018, 11:12 AM), https://www.vice.com /en/article/qvq9ep/companies-are-using-ai-to-stop-bias-in-hiring-they-could-also-make-discrimination-worse; Vasily Voropaev, *The Era of HR Is Coming to an End*, FORBES (Mar. 7, 2022, 6:15 AM), https://www.forbes.com/sites/forbestechcouncil/2022/03/07/the-era-of-hr-is-coming-to-an-end/.

[43] J. Edward Moreno, *Disability Bias in AI Hiring Tools Targeted in US Guidance (1)*, BLOOMBERG L. (May 12, 2022, 4:37 PM), https://news.bloomberglaw.com/daily-labor-report/disability-bias-in-ai-hiring-tools-targeted-in-federal-guidance.

[44] Samantha Cole, *Amazon Pulled the Plug on an AI Recruitment Tool That Was Biased Against Women*, VICE (Oct. 10, 2018, 11:49 AM), https://www.vice.com/en/article /evwkk4/amazon-ai-recruitment-hiring-tool-gender-bias.

[45] *Id.*.

[46] R. Stuart Geiger, Dominique Cope, Jamie Ip, Marsha Lotosh, Aayush Shah, Jenny Weng & Rebekah Tang, *"Garbage In Garbage Out" Revisited: What Do Machine Learning Application Papers Report About Human-Labeled Training Data?*, 2 QUANTITATIVE SCI. STUDS. 795, 795 (2021).

[47] Tatiana Walk-Morris, *These Are the Flaws of AI in Hiring and How To Tackle Them*, WORLD ECON. F. (Dec. 22, 2022), https://www.weforum.org/agenda/2022/12/ai-hiring-tackle-algorithms-employment-job/ [https://perma.cc/4VV5-Q954] ("Companies are increasingly recruiting staff using AI-based algorithms. . . .").

[48] *See* Mikael-Debass, *supra* note 42 ("[C]ompanies are embracing the theory that removing people from at least some parts of the hiring process can remove human bias.").

[49] *See* Robin Young & Serena McMahon, *Name Discrimination Study Finds Lakisha and Jamal Still Less Likely To Get Hired than Emily and Greg*, WBUR (Aug. 18, 2021), https://www.wbur.org/hereandnow/2021/08/18/name-discrimination-jobs [https://perma.cc/6J4E-FCJT] (revealing in economic study "[a]pplicants with Black names were called back 10% fewer times across the board . . . despite having comparable applications to their white counterparts" and "[a] trained HR specialist may be more likely to recognize bias or specifically look for diverse applicants").

job candidates' scores based on whether candidates have a bookshelf in the background or wear glasses or a headscarf.[50]

There are significant concerns, too, about the use of recruitment algorithms to weed out job applicants with disabilities. In May 2022, the Department of Justice ("DOJ") and the Equal Employment Opportunity Commission ("EEOC") issued guidance on how to use recruitment algorithms in compliance with the Americans with Disabilities Act ("ADA").[51] The EEOC noted that "an algorithmic decision-making tool could screen out an individual because of a disability" in violation of the ADA.[52] The EEOC's guidance cites examples such as chatbots that automatically screen out applicants with employment gaps, evaluation tools that take into account keystrokes per minute as a measurement of productivity, and gamified tests that cannot be performed by blind applicants. Such screen-outs are unlawful under the ADA if the applicant is nonetheless able to perform the essential functions of the job with reasonable accommodations.[53]

This case study of AI recruitment tools again illustrates how the use of AI systems might be construed as resulting in either harms or risks. On the one hand, individual job applicants may be harmed by AI systems and, therefore, could sue for discrimination, for example, under the ADA. On the other hand, the harms of such AI systems have systemic sources and, thus, could be mitigated before the fact. Bad training data and thoughtless technological design ultimately affect the entire system and all applicants, not just one person.

Regulators in this sector are already constructing these harms as risks. Some of the EEOC's proposed solutions to combat these harms constitute risk

[50] Elisa Harlan & Oliver Schnuck, *Objective or Biased*, BR24 (Feb. 16, 2021), https://interaktiv.br.de/ki-bewerbung/en/ [https://perma.cc/F6PF-FTUK] (explaining team of reporters uncovered these results while performing experiments on start-up AI developer promising application process that is "more objective and fair").

[51] *The Americans with Disabilities Act and the Use of Software, Algorithms, and Artificial Intelligence To Assess Job Applicants and Employees*, U.S. EQUAL EMP. OPPORTUNITY COMM'N, https://www.eeoc.gov/laws/guidance/americans-disabilities-act-and-use-software-algorithms-and-artificial-intelligence [https://perma.cc/6UT8-UXVA] (last visited Sept. 28, 2023). Credit ratings also affect hiring. The Federal Trade Commission ("FTC"), which enforces against credit discrimination, has twice issued guidance more generally about the use of automated decision making and its "potential for unfair or discriminatory outcomes or the perpetuation of existing socioeconomic disparities." Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM'N.: BUS. BLOG (Apr. 8, 2020), https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms [https://perma.cc/8WTX-58JP]; Elisa Jillson, *Aiming for Truth, Fairness, and Equity in Your Company's Use of AI*, FED. TRADE COMM'N: BUS. BLOG (Apr. 19, 2021), https://www.ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai [https://perma.cc/C4LD-SQMU].

[52] U.S. EQUAL EMP. OPPORTUNITY COMM'N, *supra* note 51.

[53] *Id.*

mitigation focused on the design of the decision-making tool.[54] Recent guidance offered by the Federal Trade Commission ("FTC"), which regulates the credit ratings and profiling used in hiring, is even more explicitly risk regulation.[55] It recommends ex ante testing and independent audits.[56]

The risk framing has its benefits: it could potentially protect against system-wide problems, changing the design of the system and catching errors through audits. However, it once again shifts discussions of discrimination and bias away from individualized, personalized harms. Instead, the risk framing has a future orientation; it looks at the aggregate impacts of the system on groups of people and tries to (often controversially)[57] quantify these harms. The risk framing avoids concerns about causality—who should be held at fault, the system developer or the employer who uses the discriminatory system and fails to offer reasonable accommodations?[58]—by aiming the fix at the AI system. And it once again assumes that we can and should continue to use these AI recruitment systems, even if they don't work well,[59] and even if they continue to cause harms.

3. AI and Public Health: Access to Care and Prescriptions

This final case study, or really case studies, echoes several of the problems identified in the previous two case studies. The below examples demonstrate, again, the potential of AI systems to produce biased and/or incorrect output by replicating biased data, relying on inaccurate proxies, or by ignoring crucial

[54] These include suggestions that employers ask the vendor that developed the algorithmic decision-making tool whether its interface was designed with individuals with disabilities in mind, or if the vendor determined whether any of the traits measured by the tool are correlated with certain disabilities. *See id.*

[55] *See* Smith, *supra* note 51 ("[E]xperience, as well as existing laws, can offer important lessons about how companies can manage the consumer protection risks of AI and algorithms."); *Jillson*, *supra* note 51 (explaining, for example, how business leaders can reduce risk of "company becoming the example of a business whose well-intentioned algorithm perpetuates racial inequity").

[56] *Jillson*, *supra* note 51 (calling testing "essential").

[57] It is very nearly impossible to neutrally measure a system's "bias," which is a contested concept. *See* Karen Hao, *This Is How AI Bias Really Happens—and Why It's So Hard To Fix*, MIT TECH. REV. (Feb. 4, 2019), https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/ [https://perma.cc/6S7Q-KJHY].

[58] Currently, the answer is the employer (by the EEOC under the ADA) and possibly the system developer (by the FTC, depending on whether it is conducting profiling that falls under the Fair Credit Reporting Act or the Equal Credit Opportunity Act, or whether it violates Section 5 of the FTC Act by making untrue promises about the technology, or engaging in otherwise unfair practices). *See* U.S. EQUAL EMP. OPPORTUNITY COMM'N, *supra* note 51; *Jillson*, *supra* note 51.

[59] Inioluwa Deborah Raji, I. Elizabeth Kumar, Aaron Horowitz & Andrew D. Selbst, *The Fallacy of AI Functionality*, FACCT '22, June 2022, at 959, 959 ("Deployed AI systems often do not work.").

context. They demonstrate, too, the problems with overreliance on a human in the loop of an AI system. This final set of case studies looks to the use of AI systems to allocate access to health care and prescriptions.

In 2019, a group of researchers published an article in *Science* on the racial bias in a risk-prediction algorithm widely used in the healthcare sector.[60] The algorithm was used by large health systems and insurers to target patients for enrollment in "'high-risk care management' programs," which seek to improve healthcare for high-risk patients by allocating additional resources to them.[61] The automated system was intended to find high-risk patients whom doctors would be prompted to consider enrolling in the program.

The researchers found that the algorithm erroneously overlooked high-risk Black patients, reducing the number of Black patients identified for the program by more than half. They found that the algorithm was not looking directly at health needs but at the proxy of health *costs*.[62] Because at a given number of chronic illnesses Black patients generated lower costs than White patients—the result of both healthcare inequities and social factors—they were being incorrectly labeled by the algorithm as having lower health risks.[63] Researchers have since expressed concerns that similar biases may affect an array of AI models being promulgated to guide clinical decision making for Covid-19.[64]

A second example is an algorithm called NarxCare that is used by doctors, pharmacies, and hospitals to identify a patient's risk of opioid abuse.[65] In a reaction to the opioid epidemic, nearly every state now maintains prescription drug databases to track prescriptions for controlled substances. NarxCare, developed by a company called Appriss, searches these state databases for red flags that might indicate "drug shopping" behavior, such as visiting multiple pharmacies or combining different prescriptions. At least eight states also use NarxCare's machine learning algorithm to assign each patient an Overdose Risk Score. That scoring algorithm goes beyond these state databases to look at records such as electronic health records and criminal justice data.

---

[60] Ziad Obermeyer, Brian Powers, Christine Vogeli & Sendhil Mullainathan, *Dissecting Racial Bias in an Algorithm Used To Manage the Health of Populations*, 366 SCIENCE 447, 447 (2019).

[61] *Id.* at 447 (explaining additional resources may include "greater attention from trained providers").

[62] *Id.* at 449.

[63] *Id.* at 450 ("[Black patients] generate lower costs than [White patients]—on average, $1801 less per year, holding constant the number of chronic illnesses (or $1144 less, if we instead hold constant the specific individual illnesses that contribute to the sum).").

[64] *See* Eliane Röösli, Brian Rice & Tina Hernandez-Boussard, *Bias at Warp Speed: How AI May Contribute to the Disparities Gap in the Time of COVID-19*, 28 J. AM. MED. INFORMATICS ASS'N 190, 190 (2021).

[65] *See* Maia Szalavitz, *The Pain Was Unbearable. So Why Did Doctors Turn Her Away?*, WIRED (Aug. 11, 2021, 6:00 AM), https://www.wired.com/story/opioid-drug-addiction-algorithm-chronic-pain/ [https://perma.cc/JA25-3YUW].

Researchers have criticized the use of this and similar screening algorithms on a number of grounds. One study noted that patients deemed "doctor-shoppers" by the algorithm are often actually cancer patients who have to see multiple specialists.[66] Another study found that patients reporting a Hispanic ethnicity and patients in urban areas are more likely to report difficulties in obtaining prescriptions.[67] Others criticize the algorithm's treatment of diagnoses of depression and PTSD as risk factors, which could have a disparate impact on women.[68] Additionally, the algorithm's use of criminal justice data could harm people of color, who are more likely to be arrested.

Then there is the problem of the disempowered human in the loop. While Appriss claims that physicians and pharmacists retain the final say on whether to prescribe, most states require doctors and pharmacists to consult state databases and screening tools or risk losing their licenses. Law enforcement also uses these databases to identify providers as "overprescribers."[69] This incentivizes providers to deny patients treatment, often with a disparate impact on already marginalized patients.[70] As Jennifer Oliva notes, the legal landscape in practice puts prescribers and distributors into a "precarious vice," with antidiscrimination law urging equal treatment on the one hand and medical liability pushing against prescriptions on the other.[71]

Once more: the harms of both systems could be characterized as individualized and concrete. Black patients don't get allocated additional healthcare resources and marginalized patients don't receive indicated pain treatment. Instead, these harms are largely constructed as risks. The FDA's approach (or as some criticize, lack of approach) to clinical decision support ("CDS") software is risk-based in nature.[72] The *Science* article calls not for stopping use of the algorithm, but, in effect, for better risk mitigation to lessen

[66] *Id.*; *see also* Michael Tang, Joseph Arthur, Akhila Reddy & Eduardo Bruera, *Deficiencies with the Use of Prescription Drug Monitoring Program in Cancer Pain Management: A Report of Two Cases*, 24 J. PALLIATIVE MED. 751, 752-53 (2021) (noting erroneous conclusions regarding opioid use "can generate conflict, jeopardize provider–patient relationship, breach trust and eventually affect the care that patients receive from their provider")).

[67] Amie Goodin, Karen Blumenschein, Patricia Rippetoe Freeman & Jeffery Talbert, *Consumer/Patient Encounters with Prescription Drug Monitoring Programs: Evidence from a Medicaid Population*, 15 PAIN PHYSICIAN (SPECIAL ISSUE) 169, 169 (2012).

[68] Szalavitz, *supra* note 65 ("[The] use of diagnostic records could have a disparate impact on women (who are more likely to suffer trauma from abuse).").

[69] Jennifer D. Oliva, *Dosing Discrimination: Regulating PDMP Risk Scores*, 110 CALIF. L. REV. 47, 105 (2022).

[70] *Id.* at 105-06.

[71] *Id.* at 108.

[72] *Id.* at 113 "([T]he FDA adopted a risk-based regulatory rubric for CDS software functions using factors from the International Medical Device Regulators Forum (IMDRF) Framework.").

bias.[73] Once again, the risk framing comes with consequences. Few are seriously calling for an end to prescription drug monitoring programs ("PDMPs") such as NarxCare even though it might not work.[74]

When compared, these three case studies indicate both divergences and patterns. Clearly, the laws and regulatory institutions addressing cars, employment, and public health differ. Arguably, the harms and values at issue differ too. But the problems—of incomplete or biased training data, poor data labeling, and poorly designed human-machine systems—resonate across sectors. And the increasingly common way of legally constructing these AI system harms is to label them as risks.

B. *The Arguments for AI Risk Regulation*

This Section outlines increasingly common arguments for using risk regulation to govern AI. In this Section, I speak from inside the box of policy discussions, outlining how these arguments typically go. In the next Part, however, I step outside of this box to talk about the consequences that come along with using risk regulation to the exclusion of other legal tools. There are good reasons to use at least some risk regulation to govern AI systems; but we must be aware of its blind spots, quirks, and shortcomings as a set of regulatory tools.

A number of scholars have called, explicitly or implicitly, for using the tools of risk regulation to govern AI systems.[75] They point out that "traditional methods of regulation . . . seem particularly unsuited to manage the risks associated with intelligent and autonomous machines."[76] The arguments for using risk regulation to govern AI systems echo arguments for using risk regulation in general.[77]

The first argument for risk regulation focuses on the obstacles to using ex post liability to govern AI. When an AI system causes harm, it can be particularly hard to determine causality ex post, because such systems are often technically

[73] Obermeyer et al., *supra* note 60, at 453 ("Because labels are the key determinant of both predictive quality and predictive bias, carful choice can allow us to enjoy the benefits of algorithmic predictions while minimizing their risks.").

[74] Oliva, *supra* note 69, at 114 ("[T]here is no scientific indication that PDMPs either reduce overdose death or improve patient outcomes.").

[75] *See* sources cited *supra* note 19.

[76] Scherer, *supra* note 19, at 356.

[77] *See* Douglas A. Kysar, *The Public Life of Private Law: Tort Law as a Risk Regulation Mechanism*, 9 EUR. J. RISK REGUL. 48, 51-55 (2018) [hereinafter Kysar, *Public Life of Private Law*] (surveying "commonly identified pros and cons of tort law as a method of regulating risk"). Regulating risk often involves society-wide tradeoffs, which might arguably be better made through a democratic institution such as Congress—and which lead individual judges to be reticent in allocating responsibility for harms in individual cases. Additionally, because of the considerable collective action problems raised by aggregate small harms, risk may more efficiently be dealt with by a central regulator ex ante than by individual dyadic litigation ex post. *See id.*

and legally opaque.[78] AI harms are often not foreseeable by a human programmer, which make them harder to disincentivize through dyadic litigation.[79] AI systems are complex systems that require technological expertise to understand, raising the costs of litigation.[80] Individuals already face access-to-justice barriers, and the barriers will be higher with respect to AI.[81]

Second, scholars relatedly argue that the nature of the AI harm make AI systems a better candidate for risk regulation than litigation. AI harms, like privacy harms and public health harms, may be latent in nature—that is, not yet vested.[82] AI harms, like privacy harms, are arguably externalities that companies do not yet have an incentive to internalize.[83] They may be hard to observe or

[78] *See* Scherer, *supra* note 19, at 371.

[79] *See id.* at 363-64. Scherer points out the unforeseeability itself may not just be foreseeable but intentional: programmers may intend to create a system that produces unforeseen decisions. Indeed, this is often touted as a benefit of AI. *Id.*

[80] *See id.* at 392 ("The theoretical expertise provided by expert witnesses is undercut by the stark reality that attorneys with sufficient resources will have little trouble locating a qualified expert who will testify in support of their position.").

[81] *See* Edwards & Veale, *supra* note 18, at 74-75 (noting access especially difficult in Europe, which generally doesn't allow class action suits).

[82] *See, e.g.*, A. Michael Froomkin, *Regulating Mass Surveillance as Privacy Pollution: Learning from Environmental Impact Statements*, 2015 U. ILL. L. REV. 1713, 1790 ("Like with the ice caps, the alternative to attempting to measure how much privacy we are destroying before it is all gone—in hopes of spurring mitigation—is not valuing privacy until it is too late do [sic] anything other than regret its loss."); Scherer, *supra* note 19, at 390 ("For emerging technologies, the reactive nature of tort law may delay the formation of industry expectations about how the courts will treat harm arising from a new technology."); Daniel J. Solove, *Privacy and Data Security Harms*, TECH., ACADS., POL'Y: BLOG (Aug. 14, 2014) https://www.techpolicy.com/Blog/August-2014/Privacy-and-Data-Security-Harms.aspx [https://perma.cc/3NPZ-TEB3] (attributing many courts' dismissal of data breach cases to general assumption that "harms must be vested—they must have already occurred" (emphasis omitted)).

[83] *See* Froomkin, *supra* note 82, at 1745-77 (proposing "Privacy Impact Notices" requirement for large data-collection projects, with goal of "creat[ing] some counter-pressure that partly internalizes the externalities, thus inducing firms to forgo the privacy-damaging programs with the lowest predicted rewards"); Omri Ben-Shahar, *Data Pollution*, 11 J. LEGAL ANALYSIS 104, 115 (2019) ("[Privacy harms] affect groups too broad and dispersed and cause injuries that are too abstract for private remedies to be effective."); Dennis D. Hirsch, *Protecting the Inner Environment: What Privacy Regulation Can Learn from Environmental Law*, 41 GA. L. REV. 1, 23 (2006) (explaining how privacy harms result from "tragedy of the commons" dynamics, similar to environmental harms).

quantify,[84] and are often grounded in politically contested concepts.[85] Governing AI harms may require making society-wide tradeoffs, for example between efficiency and fairness.[86] These kinds of harms and decisions, so the argument goes, are better addressed through the institutions of risk regulation than dyadic litigation.[87]

Third, researchers, including myself, have pointed out the benefits of using ex ante regulation to govern AI systems.[88] Using ex ante regulation sidesteps problems of causality, foreseeability, and control.[89] It potentially takes advantage of the institutional competencies of agencies, specialized regulators, or private firms themselves.[90] Ex ante regulation of AI systems has the potential to influence or even dictate system design, addressing harms on a collective level rather than waiting for individuals to invoke their rights.[91] Ex ante regulation of AI is thus precautionary in a general sense, in that it can prevent harms from happening rather than merely compensate for them.

Fourth, researchers make arguments both directly and obliquely for the benefits of risk regulation specifically. Matthew U. Scherer identifies the risks

[84] *See* Scherer, *supra* note 19, at 369-73 (explaining "discreet" and "diffuse" nature of AI development and "opaque" aspects of AI systems, and how these characteristics "complicate efforts to ensure that victims receive compensation ex post when AI systems cause harm"); Solove, *supra* note 82 ("Harms involving non-embarrassing data, however, are quite challenging to understand and also present some difficult practical issues.").

[85] *See* Barocas & Selbst, *supra* note 9, at 672 (arguing addressing algorithmic harms may require "a wholesale reexamination of the meanings of 'discrimination' and 'fairness'"); Pauline T. Kim, *supra* note 41, at 193(2017) ("What constitutes forbidden discrimination is highly contested in the legal and political spheres, and these debates pose a problem for computer programmers.").

[86] *See* Guihot et al., *supra* note 19, at 417-18 ("Those charged with developing principles will need to consider not only the technological and scientific concerns but also a range of societal norms and social and economic considerations."); Margot E. Kaminski & Jennifer M. Urban, *The Right to Contest AI*, 121 COLUM. L. REV. 1957, 1988-91 (2021) (discussing efficiency).

[87] *See* Scherer, *supra* note 19, at 376-92 (evaluating regulatory strengths and weaknesses of legislatures, agencies, and courts).

[88] *See* Kaminski, *Binary Governance*, *supra* note 27, at 1557-59; Edwards & Veale, *supra* note 18, at 74-80 (calling for emphasizing Impact Assessments instead of individual explanation right).

[89] *See* Scherer, *supra* note 19, at 373 ("The problem of control presents considerable challenges in terms of limiting the harm caused by AI systems once they have been developed, but it does not make it any more difficult to regulate or direct AI development ex ante.").

[90] *See* Kaminski, *Binary Governance*, *supra* note 27, at 1559-63 (proposing "collaborative governance" regime "that aims to pull inputs from, obtain buy-in from, and affect the internal institutional structures and decision-making heuristics of the private sector, while maintaining the legitimacy, efficacy, and public-interest orientation of public sector governance"); Scherer, *supra* note 19, at 383, 387 (discussing agency expertise and ability to act ex ante).

[91] *See* Kaminski, *Binary Governance*, *supra* note 27, at 1557-59.

of AI as analogous to other public risks and calls for risk regulation.[92] Michael Guihot, Anne F. Matthew, and Nicolas P. Suzor similarly identify AI risks, call for the categorization of AI system uses by risk level, and argue for a softer-law version of risk regulation.[93] Alicia Solow-Niederman identifies the regulation of AI as a form of risk regulation.[94] Other researchers do not argue explicitly for risk regulation but argue for using particular tools from the risk regulation toolkit. For example, Andrew Tutt has called for establishing a licensing regime under an "FDA for Algorithms."[95] Frank Pasquale and Gianclaudio Malgieri call for risk self-assessments and certifications, coupled with what is in effect an ex ante licensing system.[96] Andrew D. Selbst has focused on impact assessments, and also called for recording and reporting requirements.[97] I too have written about impact assessments, with Malgieri, as have a host of other authors.[98]

Researchers advocating for ex ante risk regulation or similar approaches do acknowledge the extensive challenges in regulating AI systems. They cite opacity, both technological and organizational.[99] They note the imbalance of technological expertise between the private sector and the government.[100] They point to the unknown unknowns.[101] While it's predictable that AI systems will fail—both because software crashes and because hybrid human-machine systems fail[102]—it's not necessarily predictable *how* they will fail. This makes establishing ex ante standards particularly challenging. Scholars note, too, the challenges raised by the emerging ecology of AI developers, companies, and users, with multiple actors and discrete steps and components, making it

[92] Scherer, *supra* note 19, at 358. Similar arguments have been made in the data privacy context, analogizing privacy harms to environmental harms. *See, e.g.*, Hirsch, *supra* note 83, at 10; Froomkin, *supra* note 82, at 1715; Ben-Shahar, *supra* note 83, at 106 ("[D]ata pollution is to our century what industrial pollution was to the last one.").

[93] Guihot et al., *supra* note 19, at 446 (evaluating applicability of "nudge theory" to AI regulation).

[94] Solow-Niederman, *supra* note 19, at 688 ("Rather than attempt to govern Al as a monolithic unit, a more prudent strategy is to target each of these inputs [for AI research and development] with an eye to making sure that the development process reflects public voices and values, ex ante.").

[95] Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83, 83 (2017).

[96] Gianclaudio Malgieri & Frank Pasquale, *From Transparency to Justification: Toward Ex Ante Accountability for AI* 10-14 (Brooklyn L. Sch., Legal Studies Paper No. 712, Brussels Priv. Hub, Working Paper No. 33, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4099657.

[97] Selbst, *Algorithmic Impact Assessments*, *supra* note 10, at 124.

[98] *See, e.g.*, Kaminski & Malgieri, *supra* note 18, at 125-26.

[99] *See* Frank PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 6-8, 51-52 (2015).

[100] Scherer, *supra* note 19, at 387; *see also* Guihot et al., *supra* note 19, at 421-22 (explaining the "Collingridge Dilemma:" technology is hard to regulate when it's emerging because of information asymmetries, but also hard to regulate once it's been widely adopted).

[101] Guihot et al., *supra* note 19, at 414-27.

[102] Crootof et al., *supra* note 8, at 468-70.

potentially hard to ascribe responsibility or identify a place of intervention for governance.[103] And they cite to the much-hyped "pacing problem," asking how to craft regulation that will keep up with technological change.[104]

In short, scholars, including myself, have been calling for using risk regulation in general, or for using specific risk regulation tools, in governing AI systems. The arguments from within the policy box are often quite convincing. But these arguments, including mine, generally fail to look at the consequences of this particular mode of legal construction.

C. *AI Risk Regulation*

Regulators have recently converged around using risk regulation to govern AI. But not all risk regulation is the same. AI risk regulation is a particular version of risk regulation with particular tools. This Section introduces risk regulation in general, and the consensus version of AI risk regulation in particular, drawing on multiple recent examples and highlighting what tools AI risk regulation deploys and omits.

1. Risk Regulation

For readers unfamiliar with risk regulation, this Subsection offers a short introduction. Risk regulation constitutes both a set of regulatory goals and a set of regulatory tools.[105] The goals are, in some sense, simple: to prevent, reduce, or mitigate significant risks, usually those arising from complex systems or technologies.

Risk regulation is often, though not always, ex ante, systemic, and concerned with aggregate outcomes.[106] It often targets a system's design and attempts to mitigate risk before harms occur. Risk regulation often involves a centralized agency, but may also entail less command-and-control governance, such as setting performance standards and delegating the details to regulated companies. While there is little consensus over the precise boundaries of risk regulation,

---

[103] *See* Scherer, *supra* note 19, at 369-73 (describing complicated barriers to AI regulation created by discreet, diffuse, discrete, and opaque research and development processes); *see also* Reuben Binns & Michael Veale, *Is That Your Final Decision? Multi-Stage Profiling, Selective Effects, and Article 22 of the GDPR*, 11 INT'L DATA PRIV. L. 319, 324-31 (2021) (discussing challenges with identifying where decision takes place).

[104] *See, e.g.*, Guihot et al., *supra* note 19, at 421; Marchant & Stevens, *supra* note 19, at 237 ("The technologies are advancing at a very rapid pace, perhaps too fast for traditional regulatory programs to keep pace.").

[105] This Article's definition of risk regulation echoes William Boyd's, but is significantly broader, as it also encompasses nonquantitative risks and includes not just risk assessment but precautionary tactics and post-market and ongoing measures. *See* William Boyd, *Genealogies of Risk: Searching for Safety, 1930s-1970s*, 39 ECOLOGY L.Q. 895, 897 (2012) ("As understood here, risk thinking is intended as shorthand for the various concepts, tools, and practices that underwrite the formal understanding and assessment of risk.").

[106] *See* Marchant & Stevens, *supra* note 19, at 251-52 (noting how current model focuses on "front-end rulemaking").

there is considerable consensus that the crux of the matter is getting the institutions right.[107]

An example or two may be helpful. A classic, oft-cited version of early risk regulation is the Delaney Clause. The Delaney Clause was added to the Federal Food Drug and Cosmetics Act in 1958.[108] It established what was, in effect, a zero-tolerance approach to any food additives that research demonstrated might be carcinogenic.[109] That is, the Delaney Clause is an early and precautionary approach to risk regulation, requiring that safety be demonstrated before a product be released.[110]

Risk regulation has since evolved to have a number of shapes and tools and institutional arrangements, from ex ante licensing, to labeling, to requiring testing before a technology is widely used.[111] For example, FDA sets tolerance levels for the use of beef hormones.[112] The EPA can, under narrow circumstances, require toxic chemical testing before chemicals are released.[113]

For purposes of this Article, it may help to think of risk regulation as having three categories of tools: precautionary tactics, risk analysis and mitigation, and post-market measures.[114]

---

[107] *See* Kysar, *Public Life of Private Law*, *supra* note 77, at 50, 64.

[108] Food Additives Amendment of 1958, Pub. L. No. 85-929, 72 Stat. 1784, 1785 (1958).

[109] *See* DAVID VOGEL, THE POLITICS OF PRECAUTION: REGULATING HEALTH, SAFETY, AND ENVIRONMENTAL RISKS IN EUROPE AND THE UNITED STATES 46 (2012); Boyd, *supra* note 105, at 901.

[110] *See* VOGEL, *supra* note 109, at 253.

[111] *See id.* at 40-42; Boyd, *supra* note 105, at 900, 904. The Toxic Substances Control Act of 1976 ("TSCA") gave the EPA authority to require testing of new chemicals and to regulate if the EPA can "demonstrate that a particular chemical posed an unreasonable risk before issuing any testing requirements and before regulating." Boyd, *supra* note 105, at 975, 977 ("[The change in language between the Delaney Clause and TSCA] also reflected a very different posture toward uncertainty and the possibility of knowledge regarding complex and emerging environmental hazards. Unreasonable risk, and the balancing that it entailed, demanded a degree of quantification and precision that was largely absent in the earlier conceptions of endangerment. There was an assumption, in other words, that risks could be quantified and understood sufficiently in order to run them through a risk-benefit analysis as a prerequisite for regulation."); Toxic Substances Control Act of 1976, Pub. L. No. 94-469, 90 Stat. 2003 (1976). The EU both uses ex ante approval (licensing) of GMOs and requires the labeling of food products containing GMOs. VOGEL, *supra* note 109, at 66-72.

[112] VOGEL, *supra* note 109, at 59.

[113] *See* sources cited *supra* note 111 and accompanying text.

[114] These categories overlap and are not exclusive; what this Article characterizes as "risk analysis and mitigation," others may characterize as "precautionary tactics," and vice versa. For example, Gary Marchant and Yvonne Stevens identify four governance tools, including "risk analysis, precaution, liability, and resilience," which they argue can be distinguished based on whether they are *ex ante* versus *ex post*, and permissive versus prohibitive." Marchant & Stevens, *supra* note 19, at 236, 245. This Article's categories differ both in placing less focus on tort liability as a distinct category of risk regulation, and in highlighting post-market measures beyond resilience.

Precautionary tactics are tactics grounded in the precautionary principle—the idea that technologies should not be used unless proven safe enough. Precautionary tactics include legal bans,[115] licensing,[116] and regulatory sandboxing.[117] In the United States, few technologies are ever actually banned.[118] Licensing is a far more common form of precautionary regulation.[119] Regulatory sandboxing, which is typically applied in highly regulated contexts, is a lighter-touch precautionary tactic that is growing in popularity, including in the context of AI governance.[120] Sandboxing permits the use of a new technology in an enforcement safe harbor, subject to regulatory supervision.

The second set of tools in the risk regulation toolkit constitutes risk assessment and mitigation. Risk assessment and mitigation typically requires developers (and less typically, users) to conduct risk analysis and mitigate risks. There can be some overlap with licensing regimes, such as when licensing is conditioned on risk mitigation or on meeting a performance standard. Some use the term "risk regulation" to refer only to these tools, and sometimes only to a very specific version of them.[121] That is, for some, risk analysis and mitigation *are* risk regulation, and the other tools in the risk regulation toolkit can get overlooked or ignored.[122]

---

[115] *See* Rebecca Crootof, *The Killer Robots Are Here: Legal and Policy Implications*, 36 CARDOZO L. REV. 1837, 1912 (2015) (discussing adoption of 1997 Mine Ban Convention, which was first time Countries agreed to ban widespread weapon).

[116] The FDA's consideration of safety and efficacy data before approving drugs is an important example of risk regulation through licensing. *See* Rachel E. Sachs, W. Nicholson Price II & Patricia J. Zettler, *Rethinking Innovation at FDA*, 104 B.U. L. REV. (forthcoming 2024) (manuscript at 53) (available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id =4373500) (arguing FDA should focus on traditional factors of safety and efficacy when making licensing decisions, instead of considering innovation incentivization).

[117] *See generally* Hilary J. Allen, *Sandbox Boundaries*, 22 VAND. J. ENT. & TECH. L. 299 (2020); Sofia Ranchordas, *Experimental Regulations for AI: Sandboxes for Morals and Mores*, 1 MORALS & MACHINES 86 (2021).

[118] *See* Crootof, *supra* note 115, at 1912.

[119] *See* Scherer, *supra* note 19, at 356.

[120] *Draft EU AI Act*, *supra* note 3, at 15 ("AI regulatory sandboxes establish a controlled environment to test innovative technologies for a limited time on the basis of a testing plan agreed with the competent authorities."). Recently, regulatory sandboxing has increased in popularity in other countries (e.g., the United Kingdom, Australia, the Netherlands, Switzerland, and Thailand), especially in the fintech space. The CFPB attempted in 2018 to establish a sandbox/safe harbor for fintech, but other regulators have contested the validity of that action. Hilary Allen notes "[i]t is therefore uncertain whether any federal regulatory sandbox is available in the United States." Allen, *supra* note 117, at 303-04 .

[121] *See* Julia Black & Robert Baldwin, *Really Responsive Risk-Based Regulation*, 32 L. & POL'Y 181, 184 (2010) (noting risk regulation frameworks vary considerably between countries).

[122] Marchant and Stevens correctly point out that "[a]n optimal governance strategy should . . . employ a mix of . . . risk governance tools, mixing and matching depending on the context and stakes." Marchant & Stevens, *supra* note 19, at 236.

The third category of risk regulation tools is post-market measures. A number of risk regulation tools are deployed after a product is already in use. These tools include revocable licenses, registration coupled with post-market monitoring, cyclical evaluation for compliance with performance metrics, and failsafe modes. Researchers have recently called for *resilience* regulation, which I believe is a form of risk regulation.[123] Rather than trying to prevent harms ex ante, resilience regulation focuses on mitigating harm when it occurs and ensuring system recovery.[124]

There is some disagreement about whether tort law is distinct from risk regulation or a part of it. Tort law on the one hand exhibits features of dyadic litigation that risk regulation seeks to avoid. On the other hand, it is well recognized that tort law can encourage regulated companies to internalize risk mitigation in the future.[125]

With all these different tools, risk regulation comes in many shapes and sizes. How policymakers approach risk can be traced to particular institutional histories, intellectual histories, broader cultures, and sociotechnical change.[126] It is all too easy to fall into the trap of thinking that there is something inevitable about how policymakers handle uncertainty. That is not so. A particular version of risk regulation is the result of particular galvanizing policy moments, particular institutional arrangements, and particular parties in power.[127]

2. AI Risk Regulation

So: which particular version of risk regulation is AI risk regulation? As discussed above, AI systems (or really, human-AI systems)[128] have certain characteristics that make them the likely targets of risk regulation. They are technologically complex. They are, at least in part, inscrutable.[129] Their use complicates debates about causality. Each of these features makes ex post litigation particularly challenging and expensive.

AI systems are also likely to fail, and to fail in unpredictable ways. All software fails at some point.[130] Moreover, coupled with their human operators,

[123] *Id.* ("While there has been some confusion in the literature about whether risk analysis is part of resilience or resilience is part of risk analysis, the two approaches are distinct but complementary.").

[124] *Id.*

[125] Kysar, *Public Life of Private Law*, *supra* note 77, at 49-52 (noting focus on ex ante deterrence as primary purpose of tort law departs from historical focus on ex post redress of wrongs).

[126] *See* Boyd, *supra* note 105, at 983 (discussing risk regulation as historically contingent, not inevitable); *see generally* Sheila Jasanoff, *The Songlines of Risk*, 8 ENV'T VALUES (SPECIAL ISSUE) 135 (1999) (discussing how culture impacts risk regulation).

[127] *See* VOGEL, *supra* note 109, at 30-40.

[128] *See, e.g.*, Selbst et al., *supra* note 7, at 60.

[129] PASQUALE, *supra* note 99, at 2 (characterizing "knowledge problem" as pervasive feature of modern economy).

[130] Choi, *supra* note 4, at 39.

AI systems constitute complex human-machine systems with well-known and likely points of failure.[131] AI systems are also arguably well suited to ex ante interventions, including design requirements that establish failure modes or facilitate accountability.[132] All of these characteristics make it unsurprising that scholars have called for risk regulation and lawmakers have deployed or proposed it.[133]

Regulators have already started to regulate the risks of AI systems. Comparing these laws and guidance from different legal regimes illustrates several key points. First, regulators are largely using risk regulation to govern AI systems. Second, there are some commonalities between these approaches, including common tools and common conceptions of risks to be mitigated. Together they provide a bigger picture of the array of tools regulators are already starting to use and the tools they are failing to deploy.

This Subsection analyzes and compares a selection of recent AI risk regulation, both proposed and enacted, in the EU and the United States. It identifies (a) what risks each law tries to address; (b) what tools each law uses to mitigate those risks; and (c) what governance style or styles each law deploys.

First, a brief word on the laws compared here. This Article draws on the examples of the EU's General Data Protection Regulation, the Draft EU AI Act, the proposed U.S. Algorithmic Accountability Act of 2022, the NIST AI Risk Management Framework, and a proposed Washington state bill. There are, to be clear, many other examples of AI risk regulation; these were chosen for analysis both because of their potential influence and because they offer a variety of different approaches to risk regulation. The GDPR has been in effect since 2018; the EU AI Act will likely become law within the next year. The NIST AI RMF is soft-law guidance currently in effect. Both the Algorithmic Accountability Act and the Washington bill have been proposed but not enacted.

The GDPR regulates the processing and use of personal data of EU persons.[134] The GDPR is at its core a data protection law (i.e., a data privacy law).[135] However, the GDPR also contains specific provisions regulating "automated individual decision-making" that produces significant effects.[136] These provisions of the GDPR regulate many, though not all, AI systems that process personal data.

---

[131] *See* Crootof et. al., *supra* note 8, at 438.

[132] Joshua A. Kroll, Joanna Huey, Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson & Harlan Yu, *Accountable Algorithms*, 165 U. PA. L. REV. 633, 696-99 (2017) (summarizing technical tools allowing decisions made by algorithms to be evaluated after the fact).

[133] *See* sources cited *supra* note 19.

[134] For more details, see Meg Leta Jones & Margot E. Kaminski, *An American's Guide to the GDPR*, 98 DEN. L. REV. 93 (2021).

[135] Thus, to the extent that the use of AI systems involves processing personal data, those uses are subject to the broader requirements of the GDPR. *See, e.g.*, Edwards & Veale, *supra* note 18, at 21; Kaminski, *Binary Governance*, *supra* note 27, at 1587-95.

[136] *See* GDPR, *supra* note 13, at art. 22.

The Draft EU AI Act was released in April 2021. The Draft EU AI Act aims to establish a comprehensive cross-sectoral European approach to governing AI systems. Once adopted and in effect, it will serve a harmonizing function, precluding EU Member States from enacting divergent laws—including more protective laws.

The NIST AI RMF is soft law—that is, voluntary rather than mandatory guidance developed through a multi-stakeholder process. In 2021, the House Appropriations Committee directed NIST to "establish a multi-stakeholder process to . . . develop a framework for managing risks related to the reliability, robustness, and trustworthiness of AI systems."[137] NIST began the process of developing its AI RMF by issuing a Request for Information in July 2021[138] and hosting a workshop in October 2021.[139] The current version of the AI RMF was finalized on January 26, 2023.[140]

In 2022, Senator Ron Wyden, Senator Cory Booker, and Representative Yvette D. Clarke introduced the Algorithmic Accountability Act.[141] The Algorithmic Accountability Act would rely on a risk mitigation approach, requiring certain companies to conduct algorithmic impact assessments and risk mitigation for the use of automated decision systems in an "augmented critical decision process."[142] The Algorithmic Accountability Act defines "critical decisions" as decisions with significant effects, including decisions relating to education, employment, essential utilities, healthcare, housing, and more.[143]

In January 2022, Washington's state legislature held a public hearing on a proposed algorithmic accountability bill, Washington Senate Bill 5116.[144] Originally introduced in 2021, the bill would govern the state government's procurement, development, and use of automated decision systems.[145]

---

[137] H.R. REP. NO. 116-455, at 23 (2021) (Conf. Rep.).

[138] *NIST Requests Information To Help Develop an AI Risk Management Framework*, NAT'L INST. STANDARDS & TECH. (July 29, 2021), https://www.nist.gov/news-events/news/2021/07/nist-requests-information-help-develop-ai-risk-management-framework [https://perma.cc/Y7TN-A2CZ].

[139] *AI Risk Management Framework Workshop Attracts 800+*, NAT'L INST. STANDARDS & TECH. (Oct. 21, 2021), https://www.nist.gov/news-events/news/2021/10/ai-risk-management-framework-workshop-attracts-800 [https://perma.cc/8GZ2-SJBH].

[140] *AI Risk Management Framework*, NAT'L INST. STANDARDS & TECH., https://www.nist.gov/itl/ai-risk-management-framework [https://perma.cc/CCE6-5EGC] (last visited Sept. 28, 2023).

[141] *See* Press Release, Ron Wyden, *supra* note 1 (noting bill updated 2019 version after input from experts and stakeholders).

[142] H.R. 6580, 117th Cong. § 3(b)(1) (2022). The Act's application is limited by company profits and the number of individuals/households impacted. *See id.* § 2 (7) (defining "covered entity").

[143] *Id.* § 2(8)(A)-(I).

[144] S.B. 5116, 67th Leg., Reg. Sess. (Wash. 2021).

[145] *Id.* § 3.

a. *The Risks*

Each of these laws constitutes AI risk regulation. But what risks, exactly, are they regulating? There is considerable overlap in what risks these laws envision: at their core most require mitigation against inaccuracy and bias. There is also, however, considerable variation in scope, with some focusing narrowly on known issues and others trying to broadly encompass any future risks from AI. Many focus, too, on risks to privacy and "AI trustworthiness," which go beyond concerns about accuracy to concerns about dignity, autonomy, accountability, and fairness.[146]

To summarize the differences: The Washington bill largely focuses on the risk of bias and discrimination.[147] The GDPR aims more broadly at protecting against risks to fundamental rights , which include but are not limited to the right to data protection.[148] The EU AI Act is concerned with two categories of risks: risks to health and safety and risks to fundamental rights.[149] The NIST AI RMF aims to mitigate an expansive variety of risks, ranging from harmful bias and poor system performance to privacy risks, cybersecurity risks, and environmental risks.[150] The Algorithmic Accountability Act similarly takes a very broad approach to contemplated risks, requiring covered companies to "[i]dentify *any likely material negative impact* . . . on consumers and assess any applicable mitigation strategy."[151]

These next paragraphs go into slightly greater detail on the risks of inaccuracy and discrimination, and the broader risks contemplated by some of these laws. They also begin to identify a core puzzle for AI risk regulation that stems from the nature of some of the risks at issue.

What each of these laws has in common is a focus on ex ante mitigation of algorithmic inaccuracy and bias. For example, the GDPR's Recitals and Guidelines build on the general foundational GDPR principles of fairness,

[146] *See, e.g.*, NIST, AI RMF, *supra* note 6, at 1 ("Understanding and managing the risks of AI systems will help to enhance trustworthiness, and in turn, cultivate public trust.").

[147] Wash. S.B. 5116 § 3.

[148] *See* Kaminski, *Binary Governance*, *supra* note 27, at 1615 ; *see, e.g.*, Quelle, *The 'Risk Revolution,' supra* note 13, at 37 ("[T]he GDPR seeks to offer a balanced form of protection of all fundamental rights that are at stake in the context of the processing of personal data.").

[149] *Draft EU AI Act*, *supra* note 3, at 13.

[150] NIST, AI RMF, *supra* note 6, at 38-39.

[151] H.R. 6580, 117th Cong. § 4(a)(9) (2022) (emphasis added).

transparency, and accuracy[152] to regulate against the risk of inaccuracy[153] and the risk of discrimination.[154] The Washington bill tasks government agencies with ensuring that government decisions made with AI do not discriminate.[155] Both the Draft EU AI Act and the proposed Algorithmic Accountability Act get granular in their requirements intended to mitigate inaccuracy and bias—with, for example, the EU AI Act setting substantive requirements for data quality and accuracy,[156] and the Algorithmic Accountability Act containing thirteen paragraphs of requirements for impact assessments, including requirements for data source documentation and ongoing system testing.[157]

Yet a number of these laws go far broader in their contemplation of the risks of AI systems. The Draft EU AI Act contemplates risks to physical safety and risks to the broad category of "fundamental rights," leaving definitions open-ended and subject to interpretation through technical standard-setting, later regulation, or interpretation by regulated private companies during

[152] GDPR, *supra* note 13, at recitals 71, 75 (outlining general risks, potential damages from risks, and best practices for processing data). The recitals are a list of reasons why a law has been adopted. Although they are non-binding, they can carry interpretive weight. They help inform the creation of the Guidelines, which detail the implementation of the GDPR. *E.g.*, Article 29 Data Prot. Working Party, Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679, 17/EN. WP 251rev.01, at 12 (Feb. 6, 2018) [hereinafter Guidelines on Automated Individual Decision-Making] (requiring data to be updated where necessary to ensure accuracy).

[153] GDPR, *supra* note 13, at recital 71 ("[T]he controller should . . . ensure, in particular, that factors which result in inaccuracies in personal data are corrected and the risk of errors is minimised . . . ."); *see also* Guidelines on Automated Individual Decision-Making, *supra* note 152, at 11-12.

[154] GDPR, *supra* note 13, at recital 71 ("[T]he controller should . . . prevent[], inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect."); *see also* Guidelines on Automated Individual Decision-Making, *supra* note 152, at 6 (pointing to "unjustified discrimination").

[155] S.B. 5116, 67th Leg., Reg. Sess. § 4 (Wash. 2021) ("A public agency may not develop, procure, or use an automated decision system that discriminates against an individual . . . .").

[156] *Draft EU AI Act*, *supra* note 3, at 7 ("For high-risk AI systems, the requirements of high quality data, documentation and traceability, transparency, human oversight, accuracy and robustness, are strictly necessary to mitigate the risks to fundamental rights and safety posed by AI . . . ."). Private organizations will conduct technical standards setting to determine the content of some of the substantive requirements of the Act. *Id.* at 13, 63; *see also* Michael Veale (@mikarv), TWITTER (Feb. 4, 2022, 2:24 AM), https://twitter.com/mikarv/status/1489500000795123715 [https://perma.cc/6AJD-UJ5Q] ("Significant news for the AI Act from the Commission as it proposes its new Standardisation Strategy, involving amending the 2012 Regulation. Remember: private bodies making standards (CEN/CENELEC/ETSI) are the key entities in the AI Act that determine the final rules."); Veale & Zuiderveen Borgesius, *supra* note 12, at 105 (criticizing standards setting). In the absence of technical standards, the European Commission may later fill in some details. *Draft EU AI Act*, *supra* note 3, at 63.

[157] H.R. 6580, 117th Cong. § 4 (2022).

implementation.[158] The GDPR, again, focuses on risks to fundamental rights writ broad, attempting to use risk mitigation to address harms to individual autonomy and dignity.[159] The GDPR similarly tasks companies with identifying risks themselves, including among possible risks: "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."[160] The NIST AI RMF, too, is explicitly open-ended, with an Appendix identifying risks to privacy, to cybersecurity, and to the environment, among others.[161] And as mentioned above, the Algorithmic Accountability Act defines risks as "any material negative impact," and tasks companies with figuring out what those risks are. [162]

Two subcategories of risks bear particular examination: risks to accountability, or what the NIST AI RMF refers to as "trustworthiness,"[163] and risks to rights that implicate individual human autonomy and dignity. These categories of risks are both odd fits for risk regulation, shoehorned in alongside more instrumental goals like accuracy.[164]

To illustrate this, I examine a taxonomy of AI risks produced during the NIST AI RMF multistakeholder drafting process. At an earlier stage of NIST's drafting process, the agency released a taxonomy of AI risks that divided AI risks into three categories: (1) technical characteristics, (2) socio-technical characteristics, and (3) guiding principles contributing to trustworthiness.[165] The first category of risks, technical characteristics, are, according to the draft, relatively straightforward and readily measured.[166] For example, developers can test for "accuracy," in the narrow sense of ensuring that a system correctly depicts a relationship in the training data by testing whether the system is

[158] *Draft EU AI Act*, *supra* note 3, at 13.

[159] Kaminski & Malgieri, *supra* note 18, at 126-27 (discussing links between GDPR's impact assessment process and its concern over harms to individual rights).

[160] GDPR, *supra* note 13, at recital 75; *see also id.* at recital 71 (stating companies have duty to safeguard personal data "in a manner that takes account of the potential risks involved for the interests and rights of the data subject").

[161] NIST, AI RMF, *supra* note 6, app. at 38-39.

[162] H.R. 6580 § 4(a)(9).

[163] NIST, AI RMF, *supra* note 6, at 12.

[164] *See* Kaminski, *Binary Governance*, *supra* note 27, at 1578; Quelle, *The 'Risk Revolution,' supra* note 13, at 56-57.

[165] NIST, AI RISK MANAGEMENT FRAMEWORK: INITIAL DRAFT 8 (2022) [hereinafter NIST, DRAFT AI RMF]; NIST, DRAFT – TAXONOMY OF AI RISK 8 (2021).

[166] NIST, DRAFT AI RMF, *supra* note 165, at 8 ("Technical characteristics in the AI RMF taxonomy refer to factors that are under the direct control of AI system designers and developers, and which may be measured using standard evaluation criteria.").

underfit or overfit to the data.[167] Or they could test for reliability,[168] robustness,[169] or resilience[170]—all technical features of a model.

But the second and third categories of risk are not so straightforward. They encompass both harms to human autonomy and dignity, and harms to accountability. These harms are contextual, often individualized, and clearly socio-technical rather than technical in nature (if there is such a thing as a purely technical risk). For example, the drafts recognize that AI systems pose risks to societal values such as privacy and antidiscrimination.[171] These risks are really risks to individual human rights that sound in autonomy and dignity. NIST acknowledges that addressing these through risk regulation at scale is challenging.[172] These risks raise issues of measurability because "[u]nlike technical characteristics, socio-technical characteristics require significant human input and cannot yet be measured through an automated process."[173]

AI systems also pose risks to broad and contingent values such as fairness, accountability, and transparency.[174] Many proposals for AI risk regulation try to address these risks by requiring, for example, explainability,[175] transparency to expert third-parties such as auditors,[176] or the insertion of a human in the loop of individual decision making.[177] But again the oddity of using ex ante systemic risk regulation to address these kinds of concerns is that such concerns aren't generally quantifiable, and sound not only on a systemic level, but in individual dignitary rights and values.

This, then, is one of the core challenges for AI risk regulation: it deploys a largely ex ante regulatory tool best suited for readily quantifiable harms to address big, often-unquantifiable, often-contested, often-contextual, and often-individualized "risks." And while the GDPR contains a complimentary set of individual rights alongside its AI risk regulation, many of the other frameworks do not. Nor do these frameworks contemplate whether there exists an adequate

[167] *Id.* at 9.

[168] Defined as "whether a model consistently generates the same results, within the bounds of acceptable statistical error." *Id.*

[169] Defined as "a measure of model sensitivity, indicating whether the model has minimum sensitivity to variations in uncontrollable factors. A robust model will continue to function despite the existence of faults in its components." *Id.* at 10.

[170] Defined as resistance to adverse attacks. *Id.*

[171] *Id.*at 10-12 (using term "managing bias" for antidiscrimination).

[172] NIST, AI RMF, *supra* note 6, at 5-6.

[173] NIST, DRAFT AI RMF, *supra* note 165.

[174] *Id.* at 13.

[175] Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 204 (2019) [hereinafter Kaminski, *Explanation, Explained*].

[176] Kaminski, *Binary Governance*, *supra* note 27, at 1568; *see also* Draft *EU Draft AI Act*, *supra* note 3, at 34.

[177] GDPR, *supra* note 13, recital 19; S.B. 5116, 67th Leg., Reg. Sess. § 4-4(a)(iv) (Wash. 2021); *see also* Crootof et al., *supra* note 8, at 480 (commenting on dignitary reasons for having human in the loop).

backstop of tort liability through which individualized harms might get addressed and remedied. This problem predictably arises as part of risk regulation's "policy baggage:" it's just not well suited to unquantifiable and contested individualized harms.

b. *The Tools*

This Subsection provides an overview of the common tools of AI risk regulation. Here's the takeaway: AI risk regulation largely deploys risk assessment and mitigation to the exclusion of other tools in the risk regulation basket. That is, with some exceptions, AI risk regulation focuses on the second category of tools discussed in Section I.C.1, leaving out precautionary measures (like bans or formal licensing) and post-market measures (like post-market monitoring and recalls). With a few exceptions, the most notable of which is the GDPR, AI risk regulation also does not provide for individual rights and individual recourse.

AI risk regulation is risk regulation "lite:" focused on impact assessments and mitigation, much of which is self-supervised and subject to ex post regulatory intervention, if any. These regimes also fail to seriously contemplate the consequences of a potential lack of civil liability as a backstop to regulatory risk regulation. In general, civil liability can be an important complement to, or component of, risk regulation.[178] But for a number of the risks contemplated in these laws, including risks to privacy and risks of discrimination, civil liability faces all sorts of hurdles, both practical and legal. These range from challenges of foreseeability and inscrutability—reasons that risk regulation is touted in this context in the first place—to issues with legal standing and trade secrets law. The likely lack of civil liability contributes to the "liteness" of AI risk regulation, in that it removes one possible sanction for those who don't take mitigation seriously. It also removes a possibly crucial feedback loop, as tort cases can contribute to the development of substantive standards in risk regulation over time.

As I've discussed at great length elsewhere, AI risk regulation is also "lite" in the sense that it uses collaborative governance, rather than top-down command-and-control regulation.[179] That is, it largely relies on the participation, or supervised self-governance, of covered entities. At first glance, European laws may appear more centralized and top-down. This is not, however, necessarily the case in practice. Both the GDPR and the EU AI Act establish or recognize centralized agencies, but also lean heavily on the tools of collaborative governance or responsive regulation: formal and informal delegation to companies, multi-stakeholder processes, certification, and standards-setting organizations. That is, these laws leave a lot for companies to figure out and implement themselves. Again, as I've identified at length elsewhere, this

[178] *See, e.g.*, Kysar, *Public Life of Private Law*, *supra* note 77, at 51.

[179] *See* Kaminski, *Binary Governance*, *supra* note 27, at 1557-77 (discussing collaborative governance); *see also* Kaminski & Malgieri, *supra* note 18, at 128-29.

delegation raises huge issues of accountability and capture.[180] The NIST AI RMF, by contrast, is just straightforwardly self-regulation. It's voluntary guidance, not top-down law. The Algorithmic Accountability Act would be somewhere in between: relying on a centralized agency, namely the FTC, to do some regulating and enforcing, but also relying heavily on organizational processes and procedures, and leaving a lot of the substantive details to companies to figure out. Of the U.S. laws, the Washington bill most directly relies on centralized governance, including involvement by the public in rulemaking to determine how to define the risks of AI.[181]

The following subsections go into greater detail on the common tools in the toolkit of AI risk regulation, and criticisms.

i. Impact Assessments

The central tool of AI risk regulation is the impact assessment. Called a "data protection impact assessment" in the GDPR, an "algorithmic impact assessment" in the Algorithmic Accountability Act, a "conformity assessment" in the EU AI Act, an "algorithmic accountability report" in the Washington bill, and "Mapping, Measuring, and Managing" in the NIST AI RMF, this tool requires entities that build or use AI systems to assess and mitigate the risks of AI and to document that process.

Confusingly, the term "impact assessment" can refer to multiple things, roughly falling into two categories. An impact assessment can consist of concrete documents that must be produced before a tool can be deployed (as in Canadian Law[182] and the Washington bill). This version of the impact assessment is more static, like a "lite" version of licensing: ex ante self-assessment as a condition of deployment. Alternatively, an impact assessment can refer to a continuous cycle of risk management, an ongoing process rather than a single static document (as in the GDPR and the NIST AI RMF).

Sometimes the lines get blurred when what appears to be a static impact assessment is coupled with other ongoing tools of risk management, like audits and third-party assessments and post-market monitoring (as in the EU AI Act). Both kinds of impact assessments are typically tools of collaborative governance in that they aim to prompt self-regulation and to change organizational culture.[183] Even the static ex ante version of impact assessments typically aims to kick off ongoing mitigation processes and organizational changes.

The following are some more detailed illustrations of what the impact assessment process might entail, drawing on the laws examined here.

---

[180] *See* Kaminski & Malgieri, *supra* note 18, at 128-29.

[181] The bill tasks the Chief Information Officer with adopting rules that, among other things, will define "systematic discrimination." S.B. 5116 § 3.

[182] *See* Selbst, *Algorithmic Impact Assessments*, *supra* note 10, at 143 (describing Canadian approach to impact assessments).

[183] *See, e.g.*, *id.* at 153; Kaminski & Malgieri, *supra* note 18, at 128; Binns, *supra* note 18, at 22.

The Washington bill is an example of the ex ante, more static, licensing-like approach to impact assessments. The Washington bill couples precautionary bans and substantive standards with an ex ante risk assessment (here called an "algorithmic accountability report") that includes a mitigation plan.[184] Unlike the other U.S. approaches, the bill starts with a set of prohibitions or bans.[185] The Washington bill's impact assessment process isn't just self-regulation; it requires a sign-off by the state's director of the office of the chief information officer ("CIO").[186] The bill's assessment process is thus more like a licensing scheme than many proposed impact assessments, in that it envisions a central regulator serving a gatekeeping function (albeit not over private companies, which aren't covered by the bill at all).

Both the Algorithmic Accountability Act and the Draft EU AI Act appear at first glance to similarly envision more static, ex ante impact assessments (or "conformity assessments" in the case of the EU AI Act). However, both use less direct ex ante regulatory oversight than the Washington bill, permitting regulated entities to, in effect, self-certify rather than going to a regulator for clearance. This makes the impact assessments in these laws less like licensing.

The Algorithmic Accountability Act is built around an impact assessment process that requires covered companies to perform an impact assessment and maintain documentation of that impact assessment.[187] It requires "each covered entity to attempt to eliminate or mitigate . . . any impact . . . that demonstrates a likely material negative impact that has legal or similarly significant effects on a consumer's life."[188] The FTC would establish more details (how detailed is unclear) as to what each impact assessment must contain and be responsible for enforcing compliance.

The Draft EU AI Act similarly requires providers of high-risk AI systems to undergo a "conformity assessment" process before releasing the AI system on the EU market.[189] The conformity assessment in effect self-certifies products for use. It is licensing "ultra-lite." It can take one of two shapes, depending on whether the risk at issue is to physical safety, or to fundamental human rights:

[184] Wash. S.B. 5116 § 5(6)(h) (requiring "description of any potential impacts of the automated decision system on civil rights and liberties and potential disparate impacts on marginalized communities, and a mitigation plan").

[185] Section 4 of the Washington bill bans public agencies from (1) "us[ing] an automated decision system that discriminates," (2) "us[ing] an automated final decision system to make a decision impacting the constitutional or legal rights . . . of any Washington resident," (3) using an "automated final decision system . . . to deploy or trigger any weapon," (4) installing, in certain public places, equipment that enables AI-enabled profiling, or (5) using "[AI]-enabled profiling to make decisions that produce legal effects or similarly significant effects concerning individuals." Wash. S.B. 5116 § 4(1)-(2). It thus more closely resembles the precautionary starting point of both the GDPR and the Draft EU AI Act.

[186] *Id.* § 5(1).

[187] H.R. 6580, 117th Cong. § 3(b)(1)(B) (2022).

[188] *Id.* § 3(b)(1)(H).

[189] *Draft EU AI Act*, *supra* note 3, at 3.

assessment by third parties or self-assessment, respectively.[190] The conformity assessment process for the class of AI systems implicating safety concerns entails inspection by third-party private entities called "notified bodies,"[191] who aim to ensure that the provider complies with the Act's requirements. These notified bodies have, at least on paper, significant fact-finding and inspection abilities. However, for the second category of high-risk AI systems, those mainly affecting fundamental rights, the conformity assessment process does not involve independent third parties, just internal self-assessment and review.[192] This more closely resembles the impact assessments in the Algorithmic Accountability Act.[193] Self-assessment is not nothing; regulators may later check a system's performance against such self-assessment. It is, however, considerably weaker than a true licensing system or even third-body certification.

The EU AI Act in particular is less static and ex ante than it might initially appear. Its approach sits somewhere between the ex ante licensing-like model of impact assessments, and the iterative, continuous model discussed below. The Act's ex ante "conformity assessments" occur against the backdrop of ongoing risk mitigation and accountability measures such as registration and post-market monitoring. The Act also requires providers of high-risk systems to establish a "risk management system . . . consist[ing] of a continuous iterative process run throughout the entire lifecycle of a high-risk AI system."[194]

This brings us to the iterative model of impact assessments, exemplified by the GDPR and the NIST AI RMF. Each of these regulatory frameworks envisions the impact assessment process as continuous, ongoing, and iterative.

The GDPR explicitly requires impact assessments for AI systems, as a type of "high risk" data-processing system.[195] The GDPR requires "an assessment of

---

[190] *Id.* at 13-4, 64.

[191] *Id.* at 60.

[192] *Id.* at 13-14, annexes at 9.

[193] In short, the providers of such systems themselves review whether the required quality management systems are in place; examine technical documentation to ensure compliance with the Act's substantive requirements; and verify that both the design and development process and post-market monitoring are consistent with technical documentation. *Id.*, annexes at 10-12.

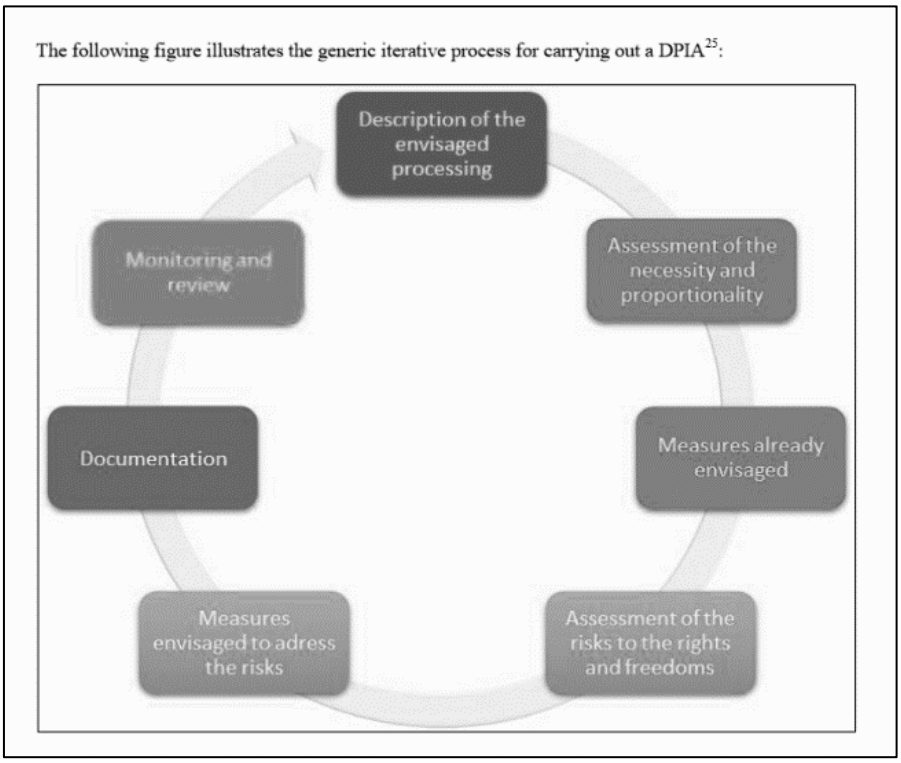
[194] *Id.* at 46. That risk management system must include (1) the identification of risks, (2) an estimation of risks when the "system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse," (3) the evaluation of other risks based on post-market monitoring; and (4) the "adoption of suitable risk management measures." *Id.* at 47. High-risk AI systems must be tested pre-market to identify appropriate risk mitigation measures. *Id.*

[195] GDPR, *supra* note 13, at art. 35(3)(a); Edwards & Veale, *supra* note 18, at 77; Casey et al., *supra* note 18, at 174. Impact assessments are required in the case of "a systematic and extensive evaluation of personal aspects relating to natural persons which is *based on automated processing*, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person." GDPR, *supra* note 13, at art. 35(3)(a) (emphasis added).

the risks to the rights and freedoms of data subjects" and a list of "measures envisaged to address the risks."[196] Companies are responsible for identifying "the origin, nature, particularity and severity of that risk."[197]

The GDPR's impact assessments are envisioned as an iterative process, rather than a one-time ex ante event.[198] A data controller is supposed to assess risks, mitigate risks, document, monitor and review, and then reengage in the process on a regular basis. This cyclical enterprise risk management is illustrated in Figure 1 *infra*.

Figure 1. GDPR Impact Assessment Process.[199]



NIST similarly envisions a cyclical, or iterative, process. The AI RMF's Core consists of three cyclical functions: Mapping, Measuring, and Managing, with a

[196] GDPR, *supra* note 13, at art. 35(7).

[197] *Id.* at recital 84.

[198] Article 29 Data Prot. Working Party, Guidelines on Data Protection Impact Assessments (DPIA) and Determining Whether Processing Is "Likely To Result in a High Risk" for the Purposes of Regulation 2016/679, 17/EN. WP 248 rev.01, at 16 n.25 (Oct. 4, 2017) [hereinafter Guidelines on DPIA].

[199] *Id.*.

Governing function that occurs throughout the process.[200] First, organizations will "Map," that is: "establish[] the context to frame risks related to an AI system."[201] Next, organizations "Measure" that is, they "analyze, assess, benchmark, and monitor AI risk and related impacts."[202] The third function consists of "Management," which might include deploying the system as is, deploying it subject to increased testing and controls, or if necessary, decommissioning the system.[203] Risk management must be ongoing.[204] The fourth Core function of the Framework, Governance, constitutes organizational risk measures designed to "implement[] a culture of risk management."[205] Governance is a "cross-cutting function" that is necessary for all the other functions to, well, function.[206]

[200] NIST, AI RMF, *supra* note 6, at 20.

[201] *Id.* at 24, 26 (noting context refers to domain and intended use, scope of system, geographical area, social environment, cultural norms, and any other specifications).

[202] *Id.* at 28.

[203] *Id.* at 31.

[204] *Id.* ("It is incumbent on Framework users to continue to apply the MANAGE function to deployed AI systems as methods, contexts, risks, and needs or expectations from relevant AI actors evolve over time.").

[205] *Id.* at 21.

[206] *Id.* at 14.

Figure 2. AI RMF Core Functions.[207]

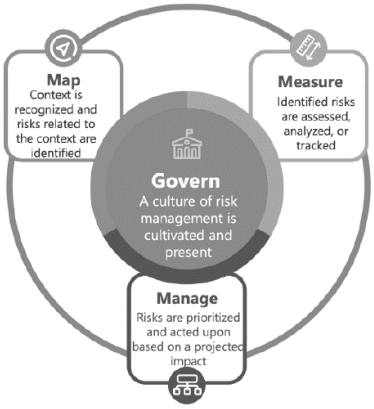


Figure 5: Functions organize AI risk management activities at their highest level to map, measure, manage, and govern AI risks. Governance is a cross-cutting function that is infused throughout and informs the other functions of the process.

The NIST AI RMF depends on a company's having an organizational culture of effective risk management.[208] This includes attributes such as accountability structures, workplace diversity, and a culture of challenging risky designs and communicating about risks.[209] NIST envisions AI risk management as just another aspect of organizational risk management that companies already deploy.[210]

Unlike ex ante impact assessments, the Framework repeatedly emphasizes risk management over the entire lifecycle of an AI system. This includes risk management at data collection, training, and post-deployment monitoring, with testing and evaluation occurring throughout the AI lifecycle:

[207] *Id.* at 20.

[208] *Id.* at 22 ("Attention to governance is a continual and intrinsic requirement for effective AI risk management over an AI system's lifespan and the organization's hierarchy.").

[209] *Id.* at 23.

[210] *Id.* at 8 ("Treating AI risks along with other critical risks, such as cybersecurity and privacy, will yield a more integrated outcome and organizational efficiencies.").

Figure 3. Stages of AI Lifecycle.[211]

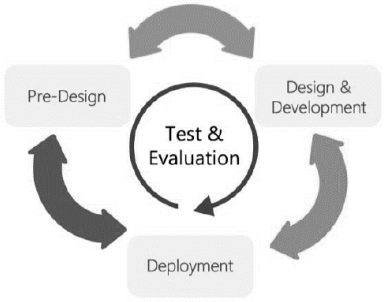


Figure 6: Risk management should be performed throughout the AI system life cycle to ensure it is continuous and timely. Example activities for each stage of the AI lifecycle follow. *Pre-Design*: data collection, curation or selection, problem formulation, and identification of stakeholders. *Design & Development*: data analysis, data cleaning, model training, and requirement analysis. *Test & Evaluation*: technical validation and verification. *Deployment*: user feedback and override, post deployment monitoring, and decommissioning.

ii. Other Tools: Audits, Testing, Precautionary Bans, Post-Market Mechanisms

This next subsection discusses the other tools used in AI risk regulation, beyond the impact assessment. As discussed in Part I.C, risk regulation can be understood to have three general buckets of tools: precautionary tactics, risk analysis and mitigation, and post-market measures. Most AI risk regulation includes additional tools beyond the impact assessment, either by reference within the impact assessment or explicitly in addition to the impact assessment. Many of these tools, such as audits, testing, and third-party oversight, constitute additional approaches to risk analysis and mitigation.[212]

It is rarer, though not unheard of, to see precautionary measures such as bans. For example, the EU AI Act bans AI systems that create an unacceptable risk, subjecting other less risky systems to risk analysis and mitigation, and to self-regulation.[213] The GDPR, too, starts from a ban of the use of solely automated

[211] NIST, DRAFT AI RMF, *supra* note 165, at 15.

[212] The GDPR Guidelines suggest that companies must assess AI risk and implement measures to try to mitigate this risk. The Guidelines recommend using expert boards, third-party and internal auditing, and developing technology to assess AI. Guidelines on Automated Individual Decision-Making, *supra* note 152, at 27-28.

[213] The Draft EU AI Act divides uses of AI systems into uses that create (1) an unacceptable risk, (2) a high risk, or (3) low or minimal risk. *Draft EU AI Act*, *supra* note 3, at 12. The Act subjects each category of risk to a different set of regulations. Title II defines and prohibits unacceptably risky uses of AI. *Id.* at 43-45. Title IX encourages low-risk uses of systems to voluntarily self-regulate with codes of conduct. *Id.* at 80. Title III regulates high-risk uses of AI systems (defined in Annex III) primarily through an ex ante conformity

decision-making with significant effects, unless risks are mitigated appropriately and in compliance with the law.[214] Unlike the other U.S. laws, the Washington bill also starts with a ban. The Washington bill would ban public agencies from using an automated decision system that discriminates, and from using an "automated final decision system" to "make a decision impacting the constitutional or legal rights . . . of any Washington resident," among other things.[215] That is, it bans serious governmental decisions conducted using AI without a human in the loop.

The Draft EU AI Act contains the biggest mix of risk regulation tools. It contains (1) precautionary bans of some AI systems; (2) risk assessment, mitigation, and post-market measures for others; and (3) voluntary self-regulation for the third class of low-risk uses of AI.[216] The core of the EU AI Act, and the aspect of greatest relevance to this Article, is its regulation of high-risk AI systems. That regulatory regime centers around the conformity assessment process described above, but also includes pre-market registration and post-market self-monitoring and reporting coupled with post-market government oversight, as in the product regulation space.[217]

There is lots to like about the EU AI Act. However, a great deal of its potential success or failure hinges on which regulators end up overseeing AI systems—for example, a regulator accustomed to regulating data processing will likely have more relevant expertise than a regulator accustomed to approving the safety

assessment process modeled on the European product safety regime. *Id.* at 46. Title III lists the uses considered "high risk," which the Commission may later expand based on a set of criteria. *Id.* at 13. Michael Veale and Frederick Zuiderveen Borgesius observe that the Act's regulation of high-risk systems relies on "1980s product safety regulation" and "standardisation bodies with no fundamental rights experience." Veale & Zuiderveen Borgesius, *supra* note 12, at 112.

[214] *See* Kaminski, *Explanation, Explained*, *supra* note 175, at 197 (discussing GDPR Article 22's right/prohibition's application only in stated circumstances). The GDPR's Article 22 (regulating "[a]utomated individual decision-making, including profiling") covers only "solely" automated systems that produce a "significant" effect on an individual. GDPR, *supra* note 13, at art. 22. Ostensibly, the GDPR bans such decisions, subject to several exceptions. If an automated decision falls into an exception, and is thus permitted, a company using the technology is still subject to a series of requirements. Article 22 does not on its face unequivocally or clearly invoke risk regulation. It instead prescribes a set of procedural rights (to contest a decision, to express one's point of view, etc.). Article 22 does, however, require "suitable measures to safeguard the data subject's rights," also known as "suitable safeguards." *Id.* This language is one hook on which the GDPR's regulation of the risks of AI hangs. These "suitable safeguards" entail both individual procedural protections and risk regulation. The GDPR Guidelines indicate that "suitable safeguards" include risk mitigation tools, both technical and regulatory. Guidelines on Automated Individual Decision-Making, *supra* note 152, at 27-28.

[215] S.B. 5116, 67th Leg., Reg. Sess. § 4(1)-(2) (Wash. 2021).

[216] *Draft EU AI Act*, *supra* note 3, at 12-14. Preemption effectively leaves the third class of systems unregulated by individual Member States. Veale & Zuiderveen Borgesius, *supra* note 12, at 109.

[217] *See* Veale & Zuiderveen Borgesius, *supra* note 12, at 111.

of elevators.[218] And, in striking contrast to the GDPR, the Draft EU AI Act contains no individual rights for those impacted by AI systems, focusing almost exclusively on systemic governance rather than on individualized harms.[219] This is odd for a law aimed in large part at protecting human rights.

iii. What AI Risk Regulation Doesn't Include

Leaving aside the EU AI Act, AI risk regulation in general underutilizes the broader tools in the risk regulation toolkit. As discussed, only three of the regulations contemplate bans on any uses of AI systems.[220] None contemplate a true licensing regime for any AI systems.[221] The closest any of these regimes come to licensing are (1) the Draft EU AI Act's third-party oversight over AI systems that are part of products that implicate safety and (2) the Washington bill's oversight over impact assessments by the state CIO.[222]

Nor do most of these laws consider conditional licensing: a potentially useful tool of risk regulation that would allow licensing for use only in certain circumstances or with promised guard rails. Conditional licensing of a sort does make an appearance, for example in the Algorithmic Accountability Act's contemplation of AI systems that will be used with "guard rails" or only in certain contexts.[223] But the Algorithmic Accountability Act's conditional licensing is self-imposed by companies, not established by a regulator.

AI risk regulation, too, lacks what liability would provide: compensation and civil recourse.[224] On the one hand, none of the laws discussed here immunize companies from liability. On the other, as mentioned above, tort law at least in the United States will not reach many of the harms we are concerned about with respect to AI systems—for all of the reasons we employ risk regulation to begin with. For example, there is no comprehensive federal data privacy law (yet?)[225], and even if there were, the Supreme Court has made it increasingly hard to find

[218] *See id.* at 106.

[219] *See* EUR. DATA PROT. BD., JOINT OPINION 5/2021 ON THE PROPOSAL FOR A REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) 8 (2021), https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf [https://perma.cc/N9Q8-XZH5] ("Whether they are end-users, simply data subjects or other persons concerned by the AI system, the absence of any reference in the text to the individual affected by the AI system appears as a blind spot in the Proposal.").

[220] The Draft EU AI Act and the Washington bill, specifically. *Draft EU AI Act*, *supra* note 3, at 12-13; Wash. S.B. 5116, § 4. Arguably, the GDPR contemplates bans too, if you count Article 22 or other de facto bans arising from other provisions. *See* Kaminski, *Explanation, Explained*, *supra* note 175, at 196-98.

[221] *See generally*, Tutt, *supra* note 95; Malgieri & Pasquale, *supra* note 96.

[222] *Draft EU AI Act*, *supra* note 3, at 60; Wash. S.B. 5116 § 5(1).

[223] H.R. 6580, 117th Cong. § 4(a)(6) (2022).

[224] Marchant & Stevens, *supra* note 19, at 242 (discussing compensation and liability); Kaminski & Urban, *supra* note 86, at 1998-99 (discussing due process values served by having right to contest AI).

[225] *See* American Data Privacy and Protection Act, H.R. 8152, 117th Cong. (2022).

standing in privacy cases.[226] Causation issues, cumulative smaller harms, a lack of physical harm: all of these features make it unlikely that tort law would work well here, at least towards providing compensation to victims of AI-related privacy harms. A few of the regulations contemplate establishing a version of individual civil recourse in the form of a right to contest AI decisions (the GDPR, the Washington bill)—something I have discussed at length elsewhere.[227] Most of them, however, do not include an individual right to contest AI.[228]

Finally, AI risk regulation typically underuses post-market tools, such as monitoring or recalls. Both the Algorithmic Accountability Act and the Washington bill largely eschew post-market measures, relying primarily on recording and transparency post-market. Other schemes do address post-market regulation of some kind: the GDPR by requiring impact assessments to be "ongoing" and by giving regulators immense inspection and information-forcing capabilities; the Draft EU AI Act by requiring post-market monitoring and contemplating recalls; even the NIST AI RMF by emphasizing that risk analysis and mitigation includes the post-deployment period. However, each of these regimes underuses the tools of resilience: required kill switches, emergency training and protocols, and establishing thresholds at which a system should get shut down.[229] Conditional licensing, discussed above, could also be a useful post-market tool, as regulators could shut down uses that fail to stay within the conditions of their license. But again as discussed above, very little AI risk regulation contemplates conditional licensing.

II. THE POLICY BAGGAGE OF RISK REGULATION

Part I identified AI risk regulation and described its current contours. This Part asks what happens once policymakers have elected to legally construct harms as risks. Regulators make three choices regarding AI systems: first, to construct harms as risks; second, to use risk regulation as a bag of legal tools; and third, to use a particular model of risk regulation. This Part examines the first two choices, while the next Part discusses different originating models of risk regulation.

AI risk regulation is a legal transplant: policymakers have been bringing law from other fields to bear on regulating AI systems. By constructing harms as risks, policymakers trigger a version of what Vanessa Casado Pérez and Yael

[226] *See* TransUnion LLC v. Ramirez, 141 S. Ct. 2190, 2204-06 (2021) (requiring historic analogue and rejecting claim for standing on basis of credit reporting agency's failure to accurately maintain its records and offer individuals their right to access and correct their information).

[227] *See generally* Kaminski & Urban, *supra* note 86.

[228] The Algorithmic Accountability Act benchmarks around a right to contest or appeal an AI decision, without requiring it. H.R. 6580 § 4(a)(8)(B). Covered entities are required to include *whether* they provide such rights in an impact assessment; they are not, however, required to provide them. *Id.* § 4(a).

[229] *See* Marchant & Stevens, *supra* note 19, at 262, 269-70.

Lifshitz refer to as natural transplants: "[c]ross-subject transplants . . . of legal rules and doctrines that occur within a jurisdiction."[230] Natural transplants typically occur when some legal decision maker, whether regulator, judge, or practitioner, imports elements of another domestic legal regime because of its perceived salience to the problem at hand.[231] Casado Pérez and Lifshitz identify several potential problems with natural transplants, focusing primarily at the potential of a poor fit to the facts.[232]

I argue that by constructing AI harms as risks, policymakers invoke, and transplant, what I call the "policy baggage" of risk regulation: the tools, tactics, and troubles already in practice in other legal fields. This is not to say that risk regulation should not be used for governing AI. In many ways, it may be a good-enough fit. But as risk regulation emerges as the dominant approach to AI governance, we need to be aware of (and correct for) the normative content of this regulatory approach, and what it isn't doing—its blind spots.

This Part draws on research on risk regulation in other fields to ask what it means to legally construct AI harms as risks, and to regulate them using risk regulation. Risk regulation often struggles with unquantifiable harms. It cloaks policy decisions in technocratic garb. Its "techno-correctionist" nature means it largely tries to fix problems with existing technologies rather than considering whether it would be better to put regulatory energy elsewhere—including not to use a technology at all.[233] Additionally, unlike tort liability, risk regulation fails to compensate victims of AI, is potentially less responsive to changing circumstances, and may be more vulnerable to capture by regulated industries.

A. *What is Risk?*

To understand what it means to legally construct AI harms as risks, we have to examine what risk is. What do policymakers mean when they talk about risks instead of harms? Risk is defined as the "possibility of loss or injury," synonymous with "peril."[234] The modern definition of "risk" reflects the concept's actuarial nature: risk is something to be measured, often mitigated, and taken into account.[235] It is tempting to view risk as a neutral or even

[230] Casado Pérez & Lifshitz, *supra* note 21, at 936. Casado Pérez and Lifshitz identify natural transplants as "a common phenomenon in different areas of law. When facing a new legal question, judges, regulators, or even private parties turn to other areas of law they are familiar with. . . . This borrowing from what is familiar translates into borrowing across subject matters." *Id.* at 937.

[231] *Id.* at 937, 939.

[232] *Id.* at 964-65 (noting potential issues of constructing wind rights based on oil and gas rights).

[233] *See* Eaglin, *supra* note 20, at 157-58.

[234] *Risk*, MERRIAM-WEBSTER'S COLLEGIATE DICTIONARY 1076 (11th ed. 2020).

[235] In fact, another definition of risk is "the chance of loss or the perils to the subject matter of an insurance contract." *Id.*

scientific concept. But as William Boyd writes, "[l]ike all concepts, risk has a distinctive genealogy, a past, a public life."[236]

According to Boyd, while the origins of the word "risk" are unknown, the modern conception of risk appears to have emerged in the sixteenth and seventeenth century.[237] Niklas Luhmann argues that risk emerged as a concept distinct from danger, chance, and fear.[238] Risk coevolved with insurance.[239] In practice, classifying harm or danger as risk typically activates a broad set of actuarial knowledge practices.[240]

Compared to closely related concepts such as danger or harm, risk typically has several unique aspects: a future orientation, an aggregate perspective, a heavy focus on rationality and quantification, causality challenges, and an element of active choice.

Risk is different from *harm*, as the legal system normally conceives of it, in that risk takes a future orientation.[241] Harm is typically *vested*.[242] Risk refers to harm that in all likelihood—or even certitude—will happen but has not happened yet. That is, we may know with certainty that a system will eventually cause harms, but not know when or how or what precise shape those harms will take. Addressing risk thus typically means dealing in these future uncertainties, using both prognostics and prevention. Uncertainties might include what kind of harms might occur, whether a harm will occur, to whom a harm will occur, and

[236] Boyd, *supra* note 105, at 898.

[237] *Id.* at 910.

[238] *Id.* (citing NIKLAS LUHMANN, RISK: A SOCIOLOGICAL THEORY 10-11 (Rhodes Barrett trans., 1993)) (explaining Luhmann finds when modern understanding of risk emerged, it was tied to outcomes of human decisions).

[239] *Id.* at 910-11 ("[W]riting about risk took off during the late eighteenth century with the significant expansion of commercial activity and various forms of insurance (marine, property, life) reflecting in part the adoption and refinement of actuarial techniques to assess risk and price insurance contracts.").

[240] *Id.* at 900 ("[T]he ongoing debate between risk and precaution cannot be viewed simply as a battle between ideas or theories but instead must be situated in a broader, more complex (and more social) terrain of knowledge practices.").

[241] For the distinction between harm and risk of harm, see, for example, standing doctrine and the Supreme Court's struggle over the justiciability of harms likely to occur in the future. Spokeo, Inc. v. Robins, 578 U.S. 330, 339 (2016) (stating injury-in-fact test injury requires plaintiff to allege injury, *i.e.*, harm, both concrete and particularized); TransUnion LLC v. Ramirez, 141 S. Ct. 2190, 2199 (2021) (holding mere risk of future harm is insufficient to constitute injury for constitutional standing purposes).

[242] Solove, *supra* note 82 ("In looking at the law, I see a general theme, which I will refer to as the 'visceral and vested approach' to harm. Harms must be *visceral*—they must involve some dimension of palpable physical injury or financial loss. And harms must be *vested*—they must have already occurred."); *see also* Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 B.U. L. REV. 793, 836 (2022) ("[C]ourts are often uncomfortable with risk, and they cling to notions of vested harm even though risk is a concept thoroughly embraced in other domains such as insurance, business, and public health, among others.").

what degree of harm might occur, from slight to catastrophic. Some of these uncertainties are measurable and manageable; others are simply unknown.[243]

Assessing risk also typically means discussing harms at the level of the collective.[244] That is, rather than preventing or compensating for individualized harms, risk thinking assesses harms at a social level.[245] It aims at the bigger picture, at populations and systems rather than at persons.

The aggregate nature of risk has several consequences. One consequence is that individual differences typically get ironed out. Risk analysis often determines acceptable risk by looking to the average man—typically, the average adult white man.[246] Anyone whose profile deviates from this benchmark—women, people of color, nonbinary people, really any minority—can suffer the consequences of policy that treats them as invisible and their needs of little worth.[247] Second, regulating risk often involves society-wide tradeoffs.[248] Even immense individual harms may get dismissed, in the face of significant collective benefits, through the lens of risk analysis. Third, because risk is typically handled in the aggregate, its regulation typically presents a collective action problem for individuals. The costs of organizing to participate in the politics of risk are often high.

Risk analysis is typically framed as highly rational. As mentioned, risk conceptually coevolved with a set of knowledge practices that includes both insurance and statistics.[249] Thus, formal risk analysis typically involves math. A mathematician defines risk using probability and loss: the likelihood of an event

---

[243] Risk management famously deals with both "known unknowns" (or events that won't certainly happen but are known to be possible risks) and "unknown unknowns" (or events that are so far outside the realm of our knowledge as to constitute a real surprise). *See* FRANK H. KNIGHT, RISK, UNCERTAINTY, AND PROFIT 19-20, 234 (1921) (identifying "unmeasurable uncertainty"); *see also* J. M. Keynes, *The General Theory of Employment*, 51 Q.J. ECON. 209, 213-14 (1937) (identifying "'uncertain' knowledge" as matters with "no scientific basis on which to form any calculable probability whatever"); Boyd, *supra* note 105, at 912 (identifying "measurable uncertainty" as "some future consequence or outcome whose probability could be calculated").

[244] Boyd, *supra* note 105, at 912.

[245] *See* Kysar, *Public Life of Private Law*, *supra* note 77, at 52 (examining judges using tort law to balance "sensitive societal tradeoffs").

[246] *See* Boyd, *supra* note 105, at 927 ("'Man' (and it was almost always adult white men who provided the basis for these averaging exercises) became an abstract 'standardized machine' in the conceptual models . . . .").

[247] *See generally* CAROLINE CRIADO PEREZ, INVISIBLE WOMEN: DATA BIAS IN A WORLD DESIGNED FOR MEN (2019) (documenting consequences of women's invisibility in statistical studies, especially in medicine).

[248] *See* Steve P. Calandrillo, *Responsible Regulation: A Sensible Cost-Benefit, Risk Versus Risk Approach to Federal Health and Safety Regulation*, 81 B.U. L. REV. 957, 996 (2001) (noting often "risk-risk tradeoff" where regulating one risk can increase a substitute risk).

[249] Boyd, *supra* note 105, at 912 ("Mean values, averages, the normal distribution—these new concepts promised to reveal a deeper social reality beyond individual variation and seemingly random events, opening up in the process new ways of being objective about human beings.").

happening multiplied by the measurable harm to be caused by that event.[250] The consequence is that quantifiable harms often take precedence over harms that are less quantifiable. Risk analysis typically entails cost-benefit analysis, particularly in the institutional context of U.S. administrative law.[251] However, not all risks are quantifiable, and not all institutions dismiss unquantifiable risks. Sometimes risk regulation can be distinctly qualitative in nature.[252] Risk can also be different degrees of *quantifiable* (versus unquantifiable) and *known* (versus unknown).[253]

Risk is frequently used to characterize harms that result from a muddled chain of causality. Some harms are more readily framed as risks by the legal system because they are particularly difficult to causally trace back to a responsible person or entity.[254] Other harms are typically characterized as risks because of a latency period between actions and harms.[255]

Finally, risk brings with it the notion of active choice or volition. *Danger* is something to be avoided at all costs; *risk* is something that we undertake in the name of benefits.[256] We don't typically opt in to being hurt. We choose, as a society, to undertake risks in the name of both present and potential societal gains.

Risk can have other relevant features too. Often, risks are externalities, meaning that absent some form of regulation, a firm may have no incentive to internalize the cost of risk mitigation. Or the term risk can be used to refer to *really, really bad* harms, causing regulators to invoke the precautionary principle.[257]

In summary, it is tempting to think that labeling harm as risk is a descriptive rather than a normative move. After all, some harms are aggregate in nature, do raise complex causal issues, and might best be dealt with ahead of time. But labeling harm as risk also constructs the problem in particular ways, invoking a specific set of legal practices and policy conflicts.

---

[250] *Id.*

[251] *Id.* at 906 ("Risk thinking has a deep affinity with consequentialist thinking, giving it a distinctive normative valence that belies its seeming neutrality.").

[252] Black & Baldwin, *supra* note 121, at 185 ("Qualitative assessments allow for more flexibility and judgment . . . .").

[253] *See* Marchant & Stevens, *supra* note 19, at 239 (noting unquantifiable risk especially important for emerging technologies).

[254] *See* Kysar, *Public Life of Private Law*, *supra* note 77, at 53 (providing climate change as difficult to trace harm because it is caused by overwhelming number of actors).

[255] *See id.*; *see also* Lujan v. Defs. of Wildlife, 504 U.S. 555, 556 (1992) (rejecting proposed theory that anyone who used any part of "contiguous ecosystem" adversely affected had standing to sue); TransUnion LLC v. Ramirez, 141 S. Ct. 2190, 2199 (2021).

[256] *See, e.g.*, Boyd, *supra* note 105, at 910, 942 (discussing shift in U.S. administrative law from protecting safety to mitigating fluid risks).

[257] *See infra* Section II.B (discussing precautionary nuclear regulation).

B. *Risk Regulation and/versus Precaution*

In this brief Subsection, I provide an overview of the precautionary principle and discuss its relationship to risk regulation. This is necessary for understanding some of the policy baggage that AI risk regulation brings.

Policymakers and scholars like to contrast what they refer to as risk regulation with what they refer to as the precautionary principle, largely by contrasting U.S. and EU law.[258] The precautionary principle is often characterized as quintessentially un-American.[259] In fact, however, some of the most robust examples of U.S. risk regulation are precautionary in nature: the Food and Drug Administration's ("FDA") regulation of medicine, for example, and the Nuclear Regulatory Commission's ("NRC") certification scheme for nuclear reactors.[260] Thus, even as I lay out the contrasts between the precautionary principle and risk regulation, it is important to understand that a broader take on risk regulation includes precautionary approaches.[261]

[258] *See, e.g.*, RICHARD A. POSNER, CATASTROPHE: RISK AND RESPONSE 140, 148-50 (2004) (describing precautionary principle as alternative to cost-benefit analysis); CASS R. SUNSTEIN, LAWS OF FEAR: BEYOND THE PRECAUTIONARY PRINCIPLE 13 (2005) (noting precautionary principle reflects idea of "[b]etter safe than sorry"); Cass R. Sunstein, *Cognition and Cost-Benefit Analysis*, 29 J. LEGAL STUD. 1059, 1068 (2000).

[259] *See* Marchant & Stevens, *supra* note 19, at 240, 242 (noting that the American approach has focused on liability); Douglas A. Kysar, *It Might Have Been: Risk, Precaution and Opportunity Cost*, 22 J. LAND USE & ENV'T L. 1, 3-4 (2006) [hereinafter Kysar, *It Might Have Been*] (noting United States favors cost-benefit analysis that predicts, weighs, and aggregates consequences of policy proposals to identify "welfare-maximizing uses of public resources," while EU approach to risk regulation is associated with precautionary principle). *But see* Jonathan B. Wiener, *Whose Precaution After All? A Comment on the Comparison and Evolution of Risk Regulatory Systems*, 13 DUKE J. COMPAR. & INT'L L. 207, 213-15 (2003) (noting common perception that EU favors precaution and United States favors more permissive regulation is oversimplified).

[260] Both regulatory schemes start from the default position of banning a technology from general use. Only later, once the technology is deemed safe enough, is its use allowed. *See* Wiener, *supra* note 259, at 227-28 n.85, 230-32. David Vogel argues that the United States was once a global leader in precautionary governance but ceded that position to the EU after an abrupt shift towards neoliberalism in the 1990s. VOGEL, *supra* note 109, at 12-13. Others contest this characterization. Jonathan B. Wiener, Brendon Swedlow, James K. Hammitt, Michael D. Rogers & Peter H. Sand, *Better Ways To Study Regulatory Elephants*, 4 EUR. J. RISK REGUL. 311, 311 (2013); *see also* Jonathan B. Wiener, *The Real Pattern of Precaution*, *in* THE REALITY OF PRECAUTION: COMPARING RISK REGULATION IN THE UNITED STATES AND EUROPE 519, 519 (Jonathan B. Wiener, Michael D. Rogers, James K. Hammitt & Peter H. Sands eds., 2011) (distinguishing "rhetoric of precaution" from "the reality of precaution").

[261] I share the view that precaution lives on in the United States, but as a "recessive strain[]" in contemporary risk regulation. Boyd, *supra* note 105, at 904; *see also* Daniel A. Farber, *Uncertainty*, 99 GEO. L.J. 901, 915 (2011) (discussing places where "[p]recaution is implicit" in U.S. laws, including Food Quality Protection Act of 1996 and its mandates around allowable pesticide levels in food for infants and children); DIDIER BOURGUIGNON, EUR. PARLIAMENTARY RSCH. SERV., THE PRECAUTIONARY PRINCIPLE: DEFINITIONS, APPLICATIONS AND GOVERNANCE 16 (2015), https://www.europarl.europa.eu/RegData/etudes/IDAN/2015/

There is no universally accepted definition of the precautionary principle.[262] The central tenet of the principle is that policymakers facing scientific uncertainty should nonetheless attempt to prevent serious harms.[263] This is actually one approach to handling risks. Regulators do not wait for harms to happen or for uncertainty to be otherwise resolved. Rather, the precautionary principle can be interpreted to justify regulatory action before cause and effect have been proven, or even to necessitate regulation until it is clear there is no danger of serious harm.[264]

The precautionary principle entails an approach to uncertainty that is often contrasted with risk regulation.[265] Nearly all of law could be characterized as risk regulation, in that a wide variety of legal tools can be used to prevent or mitigate risks.[266] As discussed, however, many scholars have a much more specific conception of risk regulation in mind. They characterize risk regulation as laws that aim to measure, mitigate, and largely accept risks, both known and unknown, as tradeoffs for economic and social benefits. A considerable number of scholars equate risk regulation even more specifically with cost-benefit analysis, where harms must be known and measured to be regulated or

573876/EPRS_IDA(2015)573876_EN.pdf [https://perma.cc/TAS9-FCP4] ("In the United States, the term 'precautionary approach' is more readily used than 'precautionary principle.'").

[262] The principle originated in German environmental law and has been incorporated in many international treaties on environmental protection. BOURGUIGNON, *supra* note 261, at 1. In 1992 the precautionary principle was codified, without definition, into EU environmental law. *See* Treaty on European Union art. 130(r), Feb. 7, 1992, 1992 O.J. (C 191). Other European institutions have applied it in other policy realms. BOURGUIGNON, *supra* note 261, at 10 (introducing UNESCO's and European Environment Agency's working definitions of precautionary principle); *see also* Case C-157/96, The Queen v. Ministry of Agric., Fisheries & Food, 1998 E.C.R. I-2236, ¶¶ 63-64 (referring to how environment policy under Treaty on the Functioning of the EU takes protective measures "based in particular on the principles that preventive action should be taken"); Case C-180/96, United Kingdom v. Comm'n, 1998 E.C.R. I-2269, ¶¶ 99-100.

[263] BOURGUIGNON, *supra* note 261, at 6 (describing central tenet of precautionary principle as "to avoid causing adverse impacts in situations of scientific uncertainty").

[264] *Id.* at 7. According to the EU General Court, "the precautionary principle can be defined as . . . requiring the competent authorities to take appropriate measures to prevent specific potential risks to public health, safety and the environment, by giving precedence to the requirements related to the protection of those interests over economic interests." Joined Cases 74, 76, 83, 85, 132, 137 & 141/00, Artegodan GmbH v. Comm'n, 2002 E.C.R. II-4948, ¶ 184.

[265] Boyd, *supra* note 105, at 898 n.5.

[266] Tort liability, for example, can be characterized as a necessary aspect of regulating risks by forcing firms to internalize and mitigate costs. *See* Marchant & Stevens, *supra* note 19, at 235-37 (enumerating risk analysis, precaution, resilience, and liability as four governance tools); Kysar, *Public Life of Private Law*, *supra* note 77, at 49 ("Over the past half-century, however, scholars influenced by legal economic theory have come to view tort law as implicitly serving a prospective, risk regulation function.").

banned.[267] This version of risk regulation assumes that in the face of uncertainty, regulators should not ban or overregulate technologies, but rather aim their efforts at lessening known and measurable harms.[268]

Risk regulation, as many scholars conceive of it, thus entails measuring and mitigating harms and avoiding unnecessarily stringent laws, while the precautionary principle emphasizes avoiding insufficiently stringent laws.[269] Ultimately, the distinction boils down to how to address uncertainty. Boyd argues that the historic shift in the United States from precautionary safety regulation to a risk assessment approach represented a shift in how to understand and characterize the limits of scientific and technical knowledge.[270] In the 1960s and 1970s, precautionary decisions by regulators "reflected a deep and longstanding concern with uncertainty."[271] Contemporary quantitative risk assessment, by contrast, can lack awareness of its own limits.

Thus, as an initial matter, policymakers drafting AI risk regulation make a normative choice not just to label AI harms as risks, but to use the legal vocabulary of risk regulation largely instead of precautionary approaches.

C. *The Policy Baggage of Risk Regulation*

Once a policymaker decides to use the legal vocabulary of risk—not just labeling harms as risks but deploying risk regulation—foreseeable consequences

---

[267] *See* sources cited *supra* note 258. Not everyone defines risk regulation in this way. Some characterize risk regulation as precautionary regulation dealing only with a narrow class of serious physical risks and harms, such as those caused by toxic chemicals or nuclear reactor meltdowns. This kind of risk regulation tends to be top-down and command-and-control. This is not the kind of risk regulation being deployed to regulate AI systems, at least not outside of safety critical contexts. *See* Roger G. Noll, *Reforming Risk Regulation,* 545 ANNALS AM. ACAD. POL. & SOC. SCI. 165, 166-67 (1996) ("[T]he term 'risk regulation' refers to a category of environmental, health, and safety issues that have four important characteristics. . . . First, the risky event is widely perceived as potentially severe in that it could cause substantial physical damage to humans or the natural environment. . . . Second . . . the risks are widely perceived to be involuntarily imposed by either nature or other people . . . Third, the nature of a risk cannot be observed by those who would suffer from it unless they exercise a degree of diligence or incur a cost that is unreasonable. . . . Fourth, actions to ameliorate the risk are likely to be costly, and identifying appropriate actions requires expertise that most citizens do not possess."); *see also* Peter Huber, *Safety and the Second Best: The Hazards of Public Risk Management in the Courts*, 85 COLUM. L. REV. 277, 312 (1985) (focusing largely on nuclear technology and environmental risks).

[268] *See* BOURGUIGNON, *supra* note 261, at 16 ("The United States has since applied a general approach whereby a hazard must be proved before public measures can be taken. . . . Although there is no fundamental difference between this concept and the precautionary principle, it clearly reflects a US preference for a pragmatic rather than a regulatory approach, and a willingness to allow companies more freedom.").

[269] VOGEL, *supra* note 109, at 17 (2012).

[270] Boyd, *supra* note 105, at 902-03.

[271] *Id.* at 902 ("[Q]uantitative risk assessment . . . carried with it important epistemic decisions about what counted as uncertainty and what sorts of knowledge claims could be made on the basis of the techniques and evidence available.").

follow. Risk regulation comes with a known set of what I call policy baggage: known tools, tactics, and accompanying troubles that policymakers, knowingly or not, have been transplanting into the regulation of AI systems.

1. Policy Baggage

Risk regulation has known limitations. This subsection covers a number of them.

Risk regulation is often inherently techno-correctionist in nature.[272] That is, it typically takes as its starting point that a technology must be fixed so that it can be used. The less precautionary risk regulation is—the fewer bans it deploys or the more it strays from centralized licensing—the more techno-correctionist it veers. This makes the nonprecautionary flavor of risk regulation that is AI risk regulation less adept at preventing truly harmful uses from happening at all.

Second, risk regulation works best on quantifiable problems, and may struggle to address normative or qualitative questions. However, few problems are in reality purely quantifiable.[273] When risk regulation is used to address problems that mix scientific and policy questions, it can obfuscate policy decisions and shield them from democratic accountability. Relatedly, risk regulation works better when there are known unknowns—not unknown unknowns, which can pop up as unwanted surprises.

Risk regulation is often not structured to make injured people whole. It typically fails to provide civil recourse for those affected by harms. Thus, when it fails to prevent harms, it can leave people uncompensated or undercompensated. This can have consequences for the dignity of affected individuals and the perceived legitimacy of regulation. Risk regulation without tort liability removes a crucial feedback loop: without civil recourse, risk regulation can become static. Attempts to make risk regulation "adaptive" or iterative can result in capture by regulated entities.

The rest of this Section goes into a number of these limitations in greater detail, providing examples from other fields.

First, risk regulation works best on quantifiable problems.[274] For example, being able to quantify and measure the acceptable levels of a toxic chemical in an environment and mitigate to an appropriate level of risk is easier than asking a developer to make sure an AI system is fair.[275] Even purportedly quantifiable problems often exist against the backdrop of scientific uncertainty (for example, how to convert the results of animal studies to human impact when studying

[272] *See* Eaglin, *supra* note 20, at 158.

[273] *See* Wendy E. Wagner, *The Science Charade in Toxic Risk Regulation*, 95 COLUM. L. REV. 1613, 1618 (1995) (discussing vagueness of science-based statutes and difficulty of translating risks into quantitative goals); Boyd, *supra* note 105, at 942 (discussing how policy choices are made to interpret scant data on potential harms).

[274] *See* Lauren E. Willis, *Performance-Based Consumer Law*, 82 U. CHI. L. REV. 1309, 1373-74 (2015) (applying performance-based regulations to quantifiable consumer transactions).

[275] *See, e.g.*, Wagner, *supra* note 273, at 1655; GDPR, *supra* note 13, at recital 71 .

carcinogens).[276] Often harms are either not quantifiable at all, or represent a mixture of quantifiable issues with hidden policy choices.[277] Where a harm is not quantifiable, risk regulation can struggle or fail to set standards and measure performance.[278]

Risk regulation's focus on quantifiable harms raises problems.[279] It can render other harder-to-measure harms invisible.[280] Ignoring nonquantifiable harms "is literally a recipe for disaster."[281] And a focus on quantifiable harms does not mean measured harms are objective. Quantification of risk raises problems around valuation, potentially trivializes future harms, excludes questions of morality or fairness, and makes nontransparent decisions about the harms it does address.[282] That is, quantitative risk assessment as a methodology is "unable in the end to account for the normativity of what the facts tell us."[283]

Where a harm mixes measurable attributes with hidden policy choices, the use of risk regulation can shield such policy choices from democratic accountability.[284] Wendy Wagner has termed this problem, when scientific or technical experts make policy decisions in the guise of technical ones, the "science charade."[285] Wagner notes that the entwinement of technical and policy decisions is often not visible to nonexperts, making it particularly hard to

[276] Boyd, *supra* note 105, at 968-69.

[277] Wagner, *supra* note 273, at 1618 (discussing mixture and balance between policy and science in realm of toxic risk).

[278] *See* Willis, *supra* note 274, at 1378-79 (discussing problematics of accurately measuring consumer data collection).

[279] *See* Farber, *supra* note 261, at 913 ("[An] over-reliance on [quantitative risk assessment] can lead to a failure to acknowledge any risks that do not lend themselves to the technique.").

[280] *See* Boyd, *supra* note 105, at 903-04; Farber, *supra* note 261, at 909 (arguing focus on conventional risk analysis can lead to "disregard of nonquantifiable risks").

[281] Farber, *supra* note 261, at 909 (noting NRC refuses to consider terrorist attacks because such risks are impossible to quantify).

[282] *See* Frank Ackerman & Lisa Heinzerling, *Pricing the Priceless: Cost-Benefit Analysis of Environmental Protection*, 150 U. PA. L. REV. 1553, 1563 (2002) (arguing cost-benefit risk assessment is neither "objective nor transparent" because of ways personal beliefs and biases impact daily decision making).

[283] Kysar, *It Might Have Been*, *supra* note 259, at 11. Take for example the conflict over how to measure carcinogens in food. Regulators struggling to determine how to read animal studies decided to use a factor of one hundred to convert to human impact. Boyd, *supra* note 105, at 932-33. This factor, while numeric, was a policy decision based on how regulators themselves understood the science. *Id.* at 933 (characterizing regulators' approach as "obviously more intuitive than scientific"). Scientists have since examined how it diverges from scientific practice. *Id.* at 933-938.

[284] *See* Wagner, *supra* note 273, at 1674 (noting mischaracterizing policy decisions "as resolvable by science results in significant obstacles to democratic participation").

[285] *Id.* at 1629.

identify and address.[286] In addition to impeding democratic participation, hybrid problems can lead to substantive harms caused by delays and inaction as technical experts search for a technical answer to what is actually a normative policy choice.[287]

A central challenge for risk regulation is what to do about the unknown.[288] On the one hand, risk regulation is arguably well suited for taking precautions to avoid wide-scale and catastrophic risks. On the other, it is exceedingly hard, if not impossible, to know, measure, and mitigate all risks in advance. This is especially true where there are unknown unknowns, including potentially catastrophic risks.[289]

Often, risk regulation does not include a compensation scheme.[290] This is in contrast with litigation, which provides remedies to make injured parties whole. This is especially the case for risk analysis and mitigation that focus only on ex ante prevention of harms—and even more true of ex ante risk regulation that is found to preclude tort liability.[291]

AI risk regulation, as discussed in Part I, largely does not include individualized ex post process. Liability, or other forms of individual process, serve important functions beyond compensation.[292] Civil recourse, among other things, respects the dignity of affected individuals. It also can enhance the perceived legitimacy of a regulatory system and make it more acceptable to those regulated by it.[293]

---

[286] *Id.* at 1628 ("[T]he esoteric nature of science-policy problems in toxic risk regulation makes it possible for these decisionmakers to blur distinctions between science and policy without the distortions being detected by most lay observers, including elected or appointed officials.").

[287] *Id.* at 1678 ("The strong correlation between agency inaction and science-based mandates is striking, with delays extending evenly across all administrations, regardless of political ideology." (footnote omitted)).

[288] *Boyd*, *supra* note 105, at 902.

[289] Farber, *supra* note 261, at 909, 958 ("Former Secretary of Defense Donald Rumsfeld famously distinguished between known knowns, known unknowns and unknown unknowns, with the latter being the most worrisome.").

[290] Marchant & Stevens, *supra* note 19, at 242.

[291] *See* Riegel v. Medtronic, Inc., 552 U.S. 312, 315 (2008) (holding medical device regulation preempted state law tort claims).

[292] *See, e.g.*, Kaminski & Urban, *supra* note 86, at 1990 ("The rationales for due process include obtaining accuracy, supporting rule of law values, and liberal theory—that is, theory that emphasizes the importance of the individual who is affected by a given decision."); John C.P. Goldberg & Benjamin C. Zipursky, *Torts as Wrongs*, 88 TEX. L. REV. 917, 918 (2010) (noting torts is not merely law of allocating costs of accidents, but "a law of wrongs and recourse").

[293] Tom R. Tyler, *Procedural Justice, Legitimacy, and the Effective Rule of Law*, 30 CRIME & JUST. 283, 283 (2003) ("Considerable evidence suggests that the key factor shaping public behavior is the fairness of the processes legal authorities use when dealing with members of the public.").

Finally, there is the question of how to keep risk regulation at pace with the objects of its regulation. Often, risk regulation is used for technologically or scientifically complex matters that change frequently over time. Liability can establish policy feedback, as the outcomes of new cases get incorporated into new regulation.[294] Absent liability, however, many (including myself) call for using techniques from "adaptive management" or "responsive regulation."[295] These techniques have their benefits, but also raise the specter of capture by regulated industry.[296]

2. The Policy Baggage of Risk Regulation Meets AI Risk Regulation

AI risk regulation currently suffers from a number of potential problems. Some of these problems reflect the inherent limits of risk regulation. Some are perhaps specific to the problem of trying to apply risk regulation to AI systems. And some reflect regulatory myopia: policymakers have repeatedly chosen to turn to the same limited set of tools, leaving out other options from the risk regulation toolkit—perhaps a function of picking a particular model of risk regulation, discussed further below. In other words, AI risk regulation bears the policy baggage of risk regulation, and the baggage of policymakers' choices to use only certain tools.

AI risk regulation aims to "fix" risky systems—and then use them. This techno-correctionist tendency misses the fact that decisions to use actuarial AI systems in the first place are political choices.[297] The choice to use risk regulation reflects a particular epistemology: the notion that such AI systems are just math, uncovering some ground truth rather than contingent social facts.[298]

---

[294] Kysar, *Public Life of Private Law*, *supra* note 77, at 63 ("[R]ather than common law litigation being displaced by more sophisticated regulatory approaches, the latter instead may well have depended on the former for their sophistication.").

[295] *See, e.g.*, Guihot et al., *supra* note 19, at 428 (noting responsive regulation "sets out a graduated pyramid of interventions by the state in policing behavior in order to encourage and direct an optimal mix of regulatory work by private and public entities"); Marchant & Stevens, *supra* note 19, at 255 ("Adaptive management is a structured, iterative process of decision making in the face of uncertainty.").

[296] *See, e.g.*, Ari Ezra Waldman, *Privacy Law's False Promise*, 97 WASH. U. L. REV. 773, 773 (2020) (arguing industry has successfully created symbols of compliance instead of real privacy protection); Ari Ezra Waldman, *Privacy, Practice, and Performance*, 110 CALIF. L. REV. 1221, 1266 (2022) ("[E]rosion of public institutional power undermines the very mechanisms that are supposed to help compliance-based governance guard against its own devolution into regulatory capture and self-regulation.").

[297] *See* Eaglin, *supra* note 20, at 163; Jessica Eaglin, *Population-Based Sentencing*, 106 CORNELL L. REV. 353, 357 (2021) ("[T]he institutionalization of actuarial risk assessments at sentencing reflects the extension of a larger, historically situated push to move judges away from passing moral judgment on individual defendants and toward basing sentencing on population-level representations of crimes and offenses.").

[298] CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY 21 (2016); Ajunwa, *Paradox of Automation*, *supra*

A second, central problem of AI risk regulation is that the risks raised by AI systems are varied, not always quantifiable, often contested, and sometimes excruciatingly or even impossibly hard to define. The regulations outlined above reflect this. Nearly all of them attempt to use risk regulation to protect against bias and discrimination, and several purport to protect other fundamental rights.[299] Even the soft-law framework offered by NIST already struggles centrally around defining and measuring the risks of AI.

Some of the harms caused by AI—physical crashes, incorrect doses of medicine, clearly erroneous decisions—may be more readily quantifiable. Many, however, are not. For example, the development and use of AI systems can cause privacy harms. Some privacy harms, such as identity theft, can be readily observed and measured. Others, such as harms to dignity or autonomy, or for the more concretely minded, exposure to *future* risks of unauthorized disclosure or identity theft, cannot.[300] As NIST initially noted, "determinations of likelihood and severity of impact of [privacy] problems are contextual and vary among cultures and individuals."[301]

AI systems force us to have epistemic humility—to acknowledge the limits of our knowledge. They are complex systems, especially so when we take into account their human developers and users.[302] Complex systems are more likely to experience unpredictable and catastrophic risks, due in part to feedback effects.[303] The distinguishing feature of the risks raised by such systems is that "numerous tiny events coexist with a few very large ones."[304] That is, we know that unlikely and potentially catastrophic events are more likely to happen with complex systems than with less complex technologies.[305] We just can't measure or predict precisely what those events will be.

note 27, at 1686 (2020); Ngozi Okidegbe, *Discredited Data*, 107 CORNELL L. REV. 2007, 2052 (2022) (describing use of algorithms as entrenching "epistemic oppression"); *see, e.g.*, Elizabeth E. Joh, *Feeding the Machine: Policing, Crime Data, and Algorithms*, 26 WM. & MARY BILL. RTS. J. 287, 292 (2017); Jessica Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 76 (2017).

[299] Specifically, the GDPR and the Draft EU AI Act, and arguably the Algorithmic Accountability Act and the Washington bill as well. *See also* ALESSANDRO MANTELERO, BEYOND DATA: HUMAN RIGHTS, ETHICAL AND SOCIAL IMPACT ASSESSMENT IN AI 171-77 (Simone van der Hof, Bibi van den Berg, Gloria González Fuster, Eva Lievens & Bendert Zevenbergen eds., 2022) (arguing risk regulation can and should be used for protecting human rights).

[300] *See* Citron & Solove, *supra* note 242, at 817.

[301] NIST, DRAFT AI RMF, *supra* note 165, at 11.

[302] *See* Crootof et al., *supra* note 8, at 467-74.

[303] Farber, *supra* note 261, at 924, 926 ("[C]omplex systems often at least approximately follow power-law distribution. . . . [F]at tails bring with them an epistemic problem.").

[304] *Id.* at 924 (quoting ALBERT-LÁSZLÓ BARABÁSI, LINKED: THE NEW SCIENCE OF NETWORKS 67-68 (1st ed. 2002)).

[305] *Id.* ("Contrasting power laws with the normal-curve governing characteristics such as human heights, a physicist who studies complex networks points out that '[i]f the heights of

AI risks, too, often involve highly contested concepts. Take discrimination. Discrimination is a contested concept in the law, with some taking an anti-classification stance (treat everyone the same) and others touting anti-subordination (recognize that treating everyone the same is not treating everyone equitably).[306] AI risk regulation frequently refers to "fairness," and determining how to measure a concept like "fairness" imports these policy clashes into risk regulation.[307] And if what we're concerned about is a fundamental harm to privacy or to dignity or autonomy, how on earth is one to put, say, this provision of the Draft EU AI Act into practice: "[t]esting shall be made against preliminarily defined metrics and probabilistic thresholds that are appropriate to the intended purpose of the high-risk AI system"?[308]

Successful risk analysis and mitigation usually involve both substantive criteria and processes.[309] AI risk regulation, however, thus far leans more heavily on process, and less on substantive criteria. In part this is because the criteria are still quickly evolving. Each regulation tries to address this in its own way: the Draft EU AI Act through delegating standards setting to private standards-setting bodies; the NIST AI RMF through acknowledging that it is not its role to set substantive risk thresholds and through making the AI Framework a "living document" that will likely reference other sources; the Algorithmic Accountability Act and the Washington bill by requiring rulemaking by a government body.

But it is not clear that kicking the criteria can down the road will result in a solution. This, once again, is because so many of the risks policymakers are trying to regulate are not solely technical problems. Instead, they are either the kind of hybrid technical-policy problems that risk regulation typically struggles with or even purely policy questions.[310] For example, by delegating much technical substance to private standards-setting bodies, the Draft EU AI Act places big, complex, and contested policy decisions in the hands of private entities.[311] Similarly, by delegating the implementation of broad terms to private

an imaginary planet's inhabitants followed a power law distribution, most creatures would be really short,' but 'nobody would be surprised to see occasionally a hundred-foot-tall monster walking down the street.'" (quoting BARABÁSI, *supra* note 304, at 67)).

[306] *See* Barocas & Selbst, *supra* note 9, at 723 (discussing two different legal approaches to address discrimination); Kim, *supra* note 41, at 193 (explaining lack of clarity around definition of discrimination makes technical solutions more complicated).

[307] Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VA. L. REV. 811, 834 (2020) (outlining two conflicting ways to measure algorithmic fairness).

[308] *Draft EU AI Act*, *supra* note 3, at 47.

[309] *See* Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 115 (2019) (suggesting adopting substantive requirements to effectively reduce risks in algorithmic decision making).

[310] Wagner, *supra* note 273, at 1618-24 (explaining limits of science-based regulation of toxic substance risks where resolution requires both scientific and policy judgments).

[311] Veale & Zuiderveen Borgesius, *supra* note 12, at 105 (arguing EU AI Act's delegation of standard-setting power to "bodies governed by private law" may be controversial).

companies, the GDPR risks letting companies define harms to "fairness" and "fundamental rights" in ways that favor their own interests.[312]

Several laws attempt to address this by requiring regulators to consult with impacted stakeholders in setting rules (e.g., the Washington bill), or by requiring companies to consult with impacted stakeholders or their representatives in the process of conducting an impact assessment (e.g., the Algorithmic Accountability Act).[313] Such consultation might bring in necessary contesting voices to risk regulation, but it also risks overburdening communities more likely to experience the negative impact of AI systems.[314]

AI risk regulation struggles with how to do omnibus regulation right, given different levels of risks posed by different uses of these systems, and by different kinds of systems. Everyone seems to agree that the harms of AI systems will be highly contextual, yet nearly every one of these regulatory schemes offers a one-size-fits-all approach to risk regulation. This approach risks under- or over-regulating different kinds of AI systems. The emerging alternative offered by the Draft EU AI Act is to place different kinds of systems into different regulatory buckets, causing potential cliff effects with tiers that are probably too sharp. Under the EU AI Act, AI systems fall either under a complete ban, require some version of risk assessment and post-market monitoring, or go unregulated entirely.

III. DIFFERENT MODELS OF RISK REGULATION

Much of this Article has been dedicated to arguing that AI risk regulation is emerging as the dominant approach to governing AI systems, and that it shares certain tools or features across jurisdictions. This final Part complicates the story. It explains that AI risk regulation, as a legal transplant, can and does have multiple possible origins. I aim to equip the reader with the ability to determine what kind of risk regulation forms the foundation of a particular law, and to understand why some flavors of risk regulation are less satisfactory to some stakeholders than others.

A legal tool with the same name appearing in different places is not necessarily the same tool.[315] Risk regulation means different things in different

[312] Kaminski, *Binary Governance*, *supra* note 27, at 1576-77.

[313] S.B. 5116, 67th Leg., Reg. Sess. § 3 (Wash. 2021); H.R. 6580, 117th Cong. § 3(b)(1)(G) (2022). The GDPR requires consultation with stakeholders "where appropriate," but the Guidelines state that it may be as simple as conducting a survey, rather than truly empowered consultation. GDPR, *supra* note 13, art. 35(9). The Guidelines do not clarify when such views must be sought, only that companies must document their reasons for not seeking input. Guidelines on DPIA, *supra* note 198, at 15.

[314] For a discussion of this tension in the environmental law context, see Jonathan Skinner-Thompson, *Procedural Environmental Justice*, 97 WASH. L. REV. 399 (2022).

[315] Observing that policymakers transplant tools is not new. *See, e.g.*, Casado Pérez & Lifshitz, *supra* note 21, at 937 . Nor is observing that taking a tool out of context gives it new or distorted meaning. *Id.* at 933 ("[P]olicymakers and courts have borrowed from other

contexts—and may envision vastly different institutional arrangements, depending on its origins. That is, the lineage of these laws matters.[316] The versions of AI risk regulation discussed here have very different origins, both across countries (unsurprisingly) and within jurisdictions (more surprisingly). Because of significant commonalities between regulatory regimes (such as the focus on impact assessments), these origins are often obscured.

There are at least four different models of risk-regulation: a highly quantitative version, a version that uses risk regulation as democratic oversight; a version focused on allocating regulatory resources by risk, and enterprise risk management. Often, policymakers do not explicitly specify which model they are pursuing. Often, too, they deploy more than one model at once. In the AI risk regulation context, this has led to recurring conflicts between stakeholders.

Some use "risk regulation" to refer to a very specific flavor of regulation. Quantitative risk assessments emerged in U.S. administrative law in the 1960s to 1980s.[317] The hallmark of that version of risk regulation is defining risk formally and quantitatively.[318] It also occurs within the U.S. administrative state. Harms must be measured to be regulated, and to do so, cost-benefit analysis is typically employed.

A second version of risk regulation uses the tools of risk regulation as democratic oversight. The National Environmental Policy Act ("NEPA") uses risk assessments for public disclosure. NEPA's Environmental Impact Statement ("EIS")[319] requires covered entities to conduct an ex ante risk assessment that is then disclosed to the public before a project is commenced. Public disclosure is not a side effect but a major purpose of the law.[320] The EIS is intended to spur public policy conversation and a "hard look" at risk factors.[321]

resource schemes, often ignoring the scientific and social differences between these natural resources."). However, to my knowledge this has not been applied in risk regulation. Nor do Casado Pérez and Lifshitz discuss the importance of the transplant framework for conversations about governance design.

316 *See* Boyd, *supra* note 105, at 898 ("[I]t is a fallacy to view risk in transhistorical terms. Like all concepts, risk has a distinctive genealogy, a past, a public life. And that past matters as we seek to understand how this particular concept and the related practices of risk assessment have come to exercise such tremendous influence over . . . institutions and activities.").

317 Boyd, *supra* note 105, at 942-48.

318 *Id.* at 897 n.2 ("The key element of formal risk thinking is calculability.").

319 *See* 42 U.S.C. § 4332(C) (2012); Kaminski & Malgieri, *supra* note 18, at 135 ("A number of US commentators have used the EIS as a model for impact assessments in other contexts."); Andrew D. Selbst, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109, 185-88 (2017) [hereinafter Selbst, *Data Policing*] (discussing ineffectiveness of EIS model); Froomkin, *supra* note 82, at 1745 (suggesting regulation of mass surveillance be "modeled on existing environmental laws, most notably [NEPA]").

320 *See* Weinberger v. Cath. Action of Haw./Peace Educ. Project, 454 U.S. 139, 143-44 (1981) (discussing balancing of NEPA's public disclosure goals with national security).

321 Farber, *supra* note 261, at 915 (discussing requirement to disclose worst-case scenario did not make it into EIS process).

A third version of risk regulation, more influential abroad, focuses on a centralized administrator who assesses risks on a macro level as it allocates regulatory resources. This version of risk regulation began in the UK, in a document referred to as the Hampton Report.[322] It then propagated internationally, including in the financial and environmental sectors.[323] The "common starting point" in these laws is that regulators "begin by identifying the risks they are seeking to manage, not the rules they have to enforce."[324] Regulators identify the risk to be managed, select a level of risk tolerance, assess the harms and the likelihood of their occurrence, assign risk scores to firms or activities (such as "high," "medium," or "low"), and link the allocation of enforcement and inspection resources to risk scores.[325] The Draft EU AI Act is a clear descendant of this kind of law. The GDPR is a more distant descendent, in that it envisions a flexible version of top-down regulation, responsive to regulated entities, and allocating regulators' enforcement resources by risk.

The fourth model of risk regulation is enterprise risk management.[326] That is, companies typically organize themselves internally to lessen their risks. They may conduct ongoing risk analysis and mitigation to avoid liability or other penalties, whether regulatory or market-based.

While enterprise risk management can occur in the absence or shadow of law, regulators can also participate by nudging companies to conduct risk mitigation through oversight, through the threat of regulatory enforcement, by offering safe harbors, or by issuing best practices or other guidance. Enterprise risk management is typically (1) cyclical and ongoing, and (2) organizational in nature. That is, unlike NEPA impact assessments or ex ante licensing, enterprise risk management continues to occur when products are deployed. And it relies centrally on the organizational culture of the companies that deploy it, including

[322] *See generally* PHILIP HAMPTON, REDUCING ADMINISTRATIVE BURDENS: EFFECTIVE INSPECTION AND ENFORCEMENT (2005).

[323] *See* Black & Baldwin, *supra* note 121, at 183.

[324] *Id.* at 184. Black and Baldwin point to the Hampton Report as an origin of this kind of regulatory style. *Id.* at 189. The Hampton Report noted, "[r]egulators, and the regulatory system as a whole, should use comprehensive risk assessment to concentrate resources on the areas that need them most; . . . [t]he few businesses that persistently break regulations should be identified quickly, and face proportionate and meaningful sanctions; . . . [and] [r]egulators should recognise that a key element of their activity will be to allow, or even encourage, economic progress and only to intervene when there is a clear case for protection." HAMPTON, *supra* note 322, at 7; *see also* Guihot et al., *supra* note 19, at 438-39 (summarizing Hampton Report).

[325] Black & Baldwin, *supra* note 121, at 184-85.

[326] KEVIN STINE, STEPHEN QUINN, GREG WITTE & R. K. GARDNER, NIST, INTEGRATING CYBERSECURITY AND ENTERPRISE RISK MANAGEMENT (ERM), at iv (2020) (noting ERM requires identifying all risks an organization faces, as well as understanding their combined likelihood and severity).

whether employees feel they can challenge management decisions, and a culture of confronting risks, including at the design stage.[327]

These four models are not intended to be exhaustive. Rather, they represent the models most frequently drawn on by the drafters of AI risk regulation. For example, the regulation of safety-critical systems such as nuclear reactors typically constitutes a very centralized, top-down approach to risk management, and is not at play here.[328]

The challenge for proponents of risk regulation is that the different versions of it—quantitative, democratic accountability, centralized allocation of enforcement resources by risk, and internal enterprise risk management—can have different goals and tactics. That is, trying to structure enterprise risk management may run counter to the democratic accountability model, because organizations will manage risk less effectively if they know they have to expose their work to the public (in effect, a risk management chilling effect). Or, an approach focused on democratic accountability may be, by necessity, more static or stochastic than the cyclical approach of enterprise risk management. (You can't run constant democratic accountability throughout a system's lifecycle, or at least it would be extremely costly. Thus most attempts at truly public accountability, like in the Washington bill, focus on ex ante democratic accountability coupled with ex post third-party or regulatory oversight.) Or, a centralized European regulator conducting a risk-based allocation of enforcement resources might not buy into quantitative cost-benefit analysis, raising conflicts with proponents of that approach. Or, a strictly quantitative approach to risk assessments may miss out on the values democratic participants care about.

For a more specific illustration of these classification conflicts in action, take the impact assessment. The fact that impact assessments have popped up all over AI regulation does not mean that all AI regulation that uses impact assessments (1) means the same thing when it uses that term, or (2) uses the tool towards the same goals.[329] The following regulations offer at least three very different origin stories for the impact assessment—and three very different conceptions of what ends the tool is meant to serve.

The first version of the impact assessment conceives of it as a public accountability measure. This version of the algorithmic impact assessment is

[327] *See id.* at 4, 35 (noting positive risk-aware culture leads to team-based approach to monitoring and managing risks).

[328] *See, e.g.*, Crootof et al., *supra* note 8, at 495 ("The NRC's 563-page Human-System Interface Design Review Guidelines, for example, provide detailed guidance for everything from interface displays and user-interface interaction to alarm systems and the design of workstations.").

[329] Andrew Selbst makes this first point—that different regulators mean different things when they use the term "impact assessment." Selbst, *Algorithmic Impact Assessments*, *supra* note 10, at 139; *see also* Kaminski & Malgieri, *supra* note 18, at 134-37 (comparing NEPA Environmental Impact Statements, Human Rights Impact Assessments, Privacy Impact Assessments, and Data Protection Impact Assessments under GDPR).

likely a descendent of NEPA's Environmental Impact Statement.[330] Remember, the EIS, which constitutes an ex ante risk assessment disclosed to the public for notice and comment, is intended to spur public debate.[331] In fact, one criticism of the EIS process is that it is all disclosure and no substance, leaving entities that comply with the process largely substantively unregulated.[332] This is the version of the impact assessment that stakeholders in debates about the regulation of AI systems frequently cite as a model.[333] It is also the version of the AI impact assessment few regulators propose, although the Algorithmic Accountability Act, with its complex public accountability processes discussed below, gets closest.

The second version of the impact assessment is likely a descendant of centralized licensing, and thus more squarely falls into the top-down version of risk regulation that allocates enforcement resources by risk. Rather than emphasizing public accountability, this version attempts to replace centralized licensing with entity-level risk analysis and mitigation, typically coupled with accountability to a centralized regulator. The EU AI Act, with its buckets and conformity assessments, exemplifies this approach. So does the GDPR's impact assessment, aimed at high-risk processing.[334]

The third version of an impact assessment is a tool of private self-governance. It comes from enterprise risk management and is a technique used by companies to internally manage risk. When NIST issues guidance on risk management, it leans heavily on this internal risk assessment process. This version of the impact assessment may not be publicly oriented, but it is ongoing and iterative in a way the democratically oriented impact assessment typically is not.

These three different origin stories of the impact assessment—as a public accountability tool, as a watered-down version of centralized licensing that allocates resources based on risk, and as enterprise risk management—lead to very different understandings of the legal tool. They lead to different features of the assessment (ex ante versus ongoing, publicly disclosed versus internal to a company). They also indicate very different goals of regulation: prioritizing public accountability versus allocating enforcement resources versus structuring internal corporate governance. There can be overlap, and synergy, between the

[330] 42 U.S.C. § 4332(C); *see also* Kaminski & Malgieri, *supra* note 18, at 135 ; Selbst, *Data Policing*, *supra* note 319, at 119.

[331] Farber, *supra* note 261, at 915.

[332] Kaminski & Malgieri, *supra* note 18, at 135 ("The EIS process is static in nature, taking place only prior to the commencement of a project. It is procedural, rather than substantive; it does not set substantive requirements, nor prohibits anybody from doing anything. And while the EIS process requires public transparency and input, it does not require ongoing monitoring for compliance." (footnotes omitted)).

[333] *See, e.g.*, Selbst, *Algorithmic Impact Assessments*, *supra* note 10, at 127 (stating approach is used at every level of government and in many different contexts, including sentencing, privacy, and surveillance).

[334] Kaminski & Malgieri, *supra* note 18, at 125.

models, too. But these different conceptions and goals often go completely unstated.

Thus, we might ask not just of impact assessments but of risk regulation more generally: what do regulators intend it to do? And what do we want it to do? Is it meant to provide public input into and oversight over significant society-wide risks? Or to function like a licensing system, with permitting and monitoring and centralized regulatory control? Or is it meant to replicate and enhance private risk-mitigation measures? Or do some of each of these things?

The same tool or set of tools can mean very different things and can be designed very differently depending on where they came from and what the historical goals of the system have been. The implication for risk regulation more generally is that genesis matters. What may look facially similar may actually be very different in both nature and goal. And conflicts between stakeholders may stem from the wish that a different version of risk regulation were being used.

AI risk regulation as a natural transplant of methods and tools from other legal fields draws on different models of risk regulation, prompting these kinds of conflicts.[335] Recurring critiques, including mine, stem from the fact that policymakers may intend to implement one model of risk regulation (e.g., enterprise risk management) while stakeholders call for another (e.g., democratic oversight).

The sources of AI risk regulation, more generally, are varied. The GDPR builds on data protection law, as do numerous regimes now copying the GDPR's toolkit. The Draft EU AI Act borrows from products regulation and the UK model of risk regulation. The NIST AI RMF builds on its framework for cybersecurity governance, which in turn draws on other models for enterprise risk management. The Washington bill draws on NEPA, and likely also data protection law. The Algorithmic Accountability Act, too, is a conscious hybrid between data protection and NEPA, with elements of enterprise risk management throughout.

AI risk regulation for the most part is not about direct democratic accountability.[336] Instead, it usually follows one of two other models: enterprise risk management or a light-touch version of centralized risk-calibrated regulation, where only the riskiest systems are subject to regulatory oversight. That is, much of what will happen in AI risk regulation will never be released outside of companies, whether to regulators, stakeholders, experts, or the

[335] Casado Pérez & Lifshitz, *supra* note 21, at 941, 951-52.

[336] The exception is the Washington Bill, which requires public comment on the draft impact assessments before they are approved, like EIS. *See* S.B. 5116, 67th Leg., Reg. Sess. § 3 (Wash. 2021) § 5(3)-(4). The Algorithmic Accountability Act does not require release of impact assessments, nor even reports to regulators, but does mandate releasing an annual report and a repository with selections from the summary reports. H.R. 6580, 117th Cong. § 6 (2022).

public.[337] Whether, absent this third-party accountability, regulated entities will truly mitigate risks on the behalf of the public remains to be seen.[338] I remain skeptical.

The argument against public transparency or even against regulatory oversight stems from instead envisioning an enterprise risk-management version of risk management. That model emphasizes harnessing internal company expertise and building an effective organizational culture, over accountability concerns.[339] Concerns about transparency or third-party oversight sound as concerns about a risk management chilling effect: such disclosure will disincentivize effective internal risk management, either because companies will fear public sanction or because they will fear the theft of corporate secrets.[340] That is not to say that accountability does not matter in the enterprise risk management model, but rather that it is understood as an issue of organizational design, not the central goal.

The Algorithmic Accountability Act offers a creative partial approach to democratic accountability, aimed at solving this purported conflict between enterprise risk management on the one hand and democratic accountability on the other. By requiring the FTC to issue an annual public report and create a publicly accessible database with elements of impact assessments, the Act tries to veer towards the democratic accountability model. But by allowing companies not to release full reports to the public, it acknowledges chilling effects concerns from proponents of enterprise risk management.[341]

This is not to say that the democratic accountability model, or the EIS model more specifically, is the right answer. One common critique, noted above, is that the EIS process is procedural rather than substantive, in contrast to say quantitative risk management, which is highly measured and often involves setting substantive risk thresholds. Another critique of the democratic accountability model is that it is static and ex ante, rather than ongoing and iterative. The Washington bill, which hews most closely to the democratic accountability model of risk management, adopts this static and ex ante approach

[337] For example, under the GDPR, companies are not legally required to release DPIAs to the public. Guidelines on DPIA, *supra* note 198, at 18 ("Publishing a DPIA is not a legal requirement of the GDPR, it is the controller´s decision to do so."). The Guidelines do attempt to establish this as a best practice, but state that companies could release partial impact assessments or even a summary. *Id.*

[338] *Id.* at 17 (explaining DPIAs aim to "manag[e] risks to the rights of the data subjects, and thus takes their perspective").

[339] Kaminski, *Binary Governance*, *supra* note 27, at 1560-61 (summarizing arguments in favor of enterprise risk-management, including companies' technical expertise and efficient work flow); *see also* Binns, *supra* note 18, at 25 (describing benefits of self-regulation, such as ability to create customized approaches); Kaminski & Malgieri, *supra* note 18, at 127 (explaining because GDPR involves company-made, internal regulations, accountability becomes important).

[340] Selbst, *Algorithmic Impact Assessments*, *supra* note 10, at 151.

[341] H.R. 6580 § 6(b)(1)(D).

to assessments, coupling it only with post-market auditing and reporting by the state CIO. By contrast, the GDPR approach, the EU AI Act approach, and the NIST AI RMF approach are all intended to be highly iterative. In these, risk management is not an ex ante checklist to be completed, but a process to be repeated and altered as risks and knowledge change.[342]

It is worth acknowledging here that human rights impact assessments do also significantly influence discussions of algorithmic impact assessments.[343] In some ways they are a good fit as a model—like algorithmic impact assessments, they address contestable and hard-to-quantify harms. But a focus on impact assessments alone leaves out the broader toolkit of risk regulation. If we are going to write the risk regulation of AI, we should be looking at the whole legal vocabulary of risk regulation.

CONCLUSION

We are at a fork in the road for AI regulation. As the regimes analyzed in Part I illustrate, the dominant mode of proposed AI regulation is risk regulation—and not just risk regulation, but a particular flavor of it, heavily dependent on internal risk assessments and mitigation and largely eschewing a wide range of other regulatory tactics. It may be simpler to tack on to familiar legal frameworks, be they cybersecurity risk management, data protection impact assessments, or, by analogy, environmental law. But before we end up trapped in path dependency, we should take a hard look at whether these choices are the right ones.[344]

Despite the problems with both risk regulation generally and AI risk regulation in particular, it is highly unlikely that AI risk regulation will get thrown out entirely. The GDPR has been in effect since 2018, cementing the impact assessment model within many organizations. There is too much momentum around the Draft EU AI Act for it to go wildly off the rails, or ultimately depart much from its risk regulation framework.[345] Civil society groups, too, have coalesced around impact assessments, albeit versions different from proposed regulation.[346] And NIST's approach to AI through the lens of enterprise risk management is consistent with its recent approach to both

[342] NIST, AI RMF, *supra* note 6, at 2-3; MANTELERO, *supra* note 299, at 173.

[343] *See* Alessandro Mantelero, *AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment*, 34 COMPUT. L. & SEC. REV. 754, 758-60 (2018); *see generally* MANTELERO, *supra* note 299.

[344] Boyd, *supra* note 105, at 904 (stating path dependency is common in the law and creates "stubborn recalcitrance to change," which may not be appropriate to address contemporary issues); *see also* Casado Pérez & Lifshitz, *supra* note 21, at 939-40 (emphasizing transplants from familiar but different legal frameworks could have tradeoffs, especially when regulating "constantly evolving new technologies").

[345] *But see* Przemysław Pałka, *Data Management Law for the 2020s: The Lost Origins and the New Needs*, 68 BUFF. L. REV. 559, 634-35 (2020) (calling instead for sectoral legislation in lieu of "one-size-fits-all" regulation for data management).

[346] *See* sources cited *supra* note 18.

cybersecurity and privacy harms.[347] Nor is it clear that abandoning the risk regulation framework is on the whole normatively desirable, given the poor fit between AI harms and other kinds of regulatory vehicles.

Thus this Article concludes with the suggestion not that we abandon AI risk regulation entirely, but that we acknowledge and address its current limitations. This Article is largely intended to be diagnostic by nature, rather than prescriptive. However, within it there are suggestions of possible alternate paths. This includes addressing known problems with risk regulation and bringing more of its wide range of existing tools to bear on regulating the risks of AI. It also entails potentially looking to solutions beyond risk regulation, including complimentary individual rights and liability.

[347] *See* NIST, FRAMEWORK FOR IMPROVING CRITICAL INFRASTRUCTURE CYBERSECURITY 3 (2018) (emphasizing covered entities have discretion in whether and how to apply the Cybersecurity Framework); NIST, PRIVACY FRAMEWORK: A TOOL FOR IMPROVING PRIVACY THROUGH ENTERPRISE RISK MANAGEMENT, VERSION 1.0 9 (2020).