THE DEVELOPING LAW OF AI: A TURN TO RISK REGULATION

Margot E. Kaminski*

April 2023

The fast-developing law of AI is shaping up to be … risk regulation. But risk regulation comes with baggage, and there is far more to the risk regulation toolkit than algorithmic impact assessments.

The focus of early conversations about the law of artificial intelligence (AI) was on the "substitution effect": What should the law do when an AI system replaces a human actor?[1] For example, what happens to tort liability if we replace a human driver with an automated car?[2] To medical liability if we replace doctors with recommendation algorithms?[3] To the legal system—and to justice—if we replace a human judge with an AI system?[4] These conversations focused, and to some extent continue to focus, on

* Associate Professor of Law, Colorado Law School; Director, Privacy Initiative, Silicon Flatirons Center; Affiliated Faculty, Information Society Project at Yale Law School.

[1] Jack M. Balkin, "The Path of Robotics Law," *California Law Review Circuit* 6 (2015): 45, 57–58; Ryan Calo, "Robots in American Law," *University of Washington School of Law Research Paper* No. 2016-04, March 15, 2016, at 5, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2737598.

[2] See, e.g., Tracy Hresko Pearl, "Hands Off the Wheel: The Role of Law in the Coming Extinction of Human-Driven Vehicles," *Harvard Journal of Law & Technology* 33 (2020): 427; Bryant Walker Smith, "Automated Driving and Product Liability," *Michigan State Law Review* (2017): 1; Matthew Wansley, "The End of Accidents," *U.C. Davis Law Review* 55 (2021): 269.

[3] A. Michael Froomkin, Ian Kerr, & Joelle Pineau, "When AIs Outperform Doctors, Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning," *Arizona Law Review* 61 (2019): 33.

[4] Betsy Cooper, "Judges in Jeopardy! Could IBM's Watson Beat Courts at Their Own Game?" *Yale Law Journal* 121 (2011): 87; Rebecca Crootof, "'Cyborg Justice' and the Risk of Technological-Legal Lock-in,"

what human actors versus machines each bring to the table,[5] how to assign responsibility,[6] and whether AI systems should themselves be treated as legal entities.[7]

Regulators, however, have taken a markedly different turn. The newly developing "law of AI"—including the European Union's massive attempt to be the first mover in the field, the EU AI Act—isn't aimed at substitution. Rather, its focus is *risk regulation*.[8]

Risk regulation is regulation that aims to mitigate risks. It often is overseen and enforced by an expert agency rather than by courts and generally aims to encourage benefits and minimize harms at the collective level rather than afford restitution or recourse at an individual level. Risk regulation is future-oriented, trying to channel technological development and uses as they occur rather than responding to harms after the fact. And risk regulation typically adopts the normative stance of opting in to a technology and its uses, assuming that the technology can, and should, be fixed so that we can use it. That is, we choose to *take* risks (albeit preferably minimized ones); we choose, by contrast, to *avoid* harms.[9]

Risk regulation has a complex and interesting history.[10] It has been used across a variety of fields, from food and drug regulation to environmental protection to governance of the financial system. It takes many forms, from highly precautionary centralized licensing systems to light-touch self-regulation. As this paper illustrates, the version of risk regulation being deployed to govern AI systems is of the light-

---

Columbia Law Review Forum 119 (2019): 233; Richard M. Re & Alicia Solow-Niederman, "Developing Artificially Intelligent Justice," *Stanford Technology Law Review* 22 (2019): 242.

[5] Rebecca Crootof, Margot E. Kaminski, & W. Nicholson Price II, "Humans in the Loop," *Vanderbilt Law Review* 76 (2023): 429.

[6] Bryan H. Choi, "Crashworthy Code," *Washington Law Review* 94 (2019): 39; Bryan H. Choi, "Software as a Profession," *Harvard Journal of Law & Technology* 33 (2020): 557; Andrew D. Selbst, "Negligence and AI's Human Users," *Boston University Law Review* 100 (2020): 1315.

[7] Samir Chopra & Laurence F. White, *A Legal Theory for Autonomous Artificial Agents* (University of Michigan Press, 2011).

[8] See generally, Margot E. Kaminski, "Regulating the Risks of AI," *Boston University Law Review* (forthcoming 2023).

[9] William Boyd, "Genealogies of Risk: Searching for Safety, 1930s–1970s," *Ecology Law Quarterly* 39 (2012): 895, 910, 942.

[10] Id.; David Vogel, *The Politics of Precaution: Regulating Health, Safety, and Environmental Risks in Europe and the United States* (Princeton University Press, 2012).

touch and bottom-up variety. And it focuses heavily on a particular regulatory instrument: the impact assessment.

As scholars have argued, in many ways risk regulation seems like a good fit for regulating the development and growing uses of AI systems.[11] AI harms tend to be systemic, occur at scale, raise causality challenges for potential litigators, and may not yet be vested (that is, they may constitute risks of future harm rather than current harm)—all challenges for liability and arguments for regulation.

But risk regulation also comes with what I call "policy baggage": known problems that have emerged in other fields.[12] Choosing to use risk regulation itself entails making a significant normative choice: to develop and use AI systems in the first place rather than adopt more precautionary approaches to AI. Using risk regulation presumes that the technology need only be tweaked at the edges. The deontological harms raised by the use of AI systems—to autonomy, dignity, privacy, equality, and other human rights—are not inherently well-suited to a risk regulation framework.[13]

In this paper I first summarize the motivation for AI risk regulation—the known harms caused by the use of AI systems. I then offer several examples of laws, both proposed and enacted, that regulate the risks of AI systems, aiming to mitigate these harms. I discuss risk regulation's policy baggage, including deep epistemological challenges and a struggle to address hard-to-quantify harms. Specifically, I argue that risk regulation embodies what Jessica Eaglin has termed a "techno-correctionist" tendency prevalent in scholarship on AI systems: the tendency to try to make technology "better" rather than to question the politics and appropriateness of its usage and to explore more systematically whether, given its harms, it should be used at all.[14] I conclude with policy suggestions, including that regulators broaden their regulatory toolkit and move away from, or at least add to, the current narrow focus on AI impact assessments. If regulators want to truly address the harms caused by AI systems, they are going to have to do better than light-touch risk regulation.

---

[11] Matthew U. Scherer, "Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies," *Harvard Journal of Law & Technology* 29 (2016): 353, 356; Alicia Solow-Niedermann, "Administering Artificial Intelligence" *Southern California Law Review* 93 (2020): 633, 653; Michael Guihot, Anne F. Matthew, & Nicolas P. Suzor, "Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence," *Vanderbilt Journal of Entertainment & Technology Law* 20 (2020): 385; Gary E. Marchant & Yvonne A. Stevens, "Resilience: A New Tool in the Risk Governance Toolbox for Emerging Technologies," *U.C. Davis Law Review* 51 (2017): 233.

[12] See Kaminski, "Regulating the Risks of AI," supra note 8 at 21.

[13] For the argument that risk regulation can and should be used for protecting human rights, see Alessandro Mantelero, *Beyond Data: Human Rights, Ethical and Social Impact Assessment in AI* (Springer, 2022).

[14] Jessica M. Eaglin, "When Critical Race Theory Enters the Law & Technology Frame," *Michigan Journal of Race and Law* 26 (2021): 151, 155.

THE HARMS OF AI

First, what is AI?[15] Different laws define AI with different degrees of breadth or specificity. At a high level, an AI system is a computer program that, as the Singapore Model AI Governance Framework describes, "seek[s] to simulate human traits such as knowledge, reasoning, problem solving, perception, learning and planning."[16] The 2021 draft version of the proposed AI Act in the European Union, in an attempt to future-proof the law against new technologies and new uses, defined AI broadly as "software that is developed with one or more of … a) machine learning approaches, (b) logic- and knowledge-based approaches, and (c) statistical approaches" and that "can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with."[17] More recently, the EU Parliament proposed amending the AI Act's definition to largely match the definition of AI used by the Organisation for Economic Cooperation and Development (OECD): 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."."[18]

Several regulatory sources characterize AI systems as computer programs capable of producing outputs with fairly minimal human involvement.[19] Machine-learning AI systems typically do so by scanning large data sets and computationally deriving methods for making future decisions based on both the

---

[15] Bryan Casey & Mark A. Lemley, "You Might Be a Robot," *Cornell Law Review* 105 (2020): 287.

[16] Model AI Governance Framework (2nd ed.), https://perma.cc/A2PR-GN7H.

[17] European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, at Title I, Article 3(1), Annex I, COM (2021) 206 final (April 21, 2021) [hereinafter Draft EU AI Act].

[18] Luca Bertuzzi, "EU Lawmakers Set to Settle on OECD Definition for Artificial Intelligence," Euractiv, Mar. 7, 2023. See also Luca Bertuzzi, "AI Act: All the Open Political Questions in the European Parliament," Euractiv, Feb. 15, 2023 (explaining that EU lawmakers earlier proposed using the definition from the U.S. National Institute of Standards and Technology (NIST)).

[19] Draft EU AI Act, at 11 ("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"). See also AI Risk Management Framework 1.0, Jan. 26, 2023, https://perma.cc/6E3K-WCXL (hereinafter NIST AI RMF 1.0), 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. AI systems are designed to operate with varying levels of autonomy (Adapted from: OECD Recommendation on AI:2019; ISO/IEC 22989:2022)").

information in the data sets and the parameters that have been set by human programmers.[20] The outputs of AI systems can be digital (as in search engine results) or very physical (as in steering a warship or a robot).

Trying to write a general "law of AI" is somewhat like trying to write a general-purpose privacy law: challenging because different uses of AI systems, like different uses of personal data, occur in wildly differing social contexts and can have wildly differing effects. For example, using an AI system to diagnose disease could involve skilled professionals (doctors and nurses) and could result in a patient's health improving or worsening. Using an AI system to drive a car typically involves a less-than-expert human driver and could result in safer driving or in a car crash. Using an AI system to predict grades for students could result in changes to students' educational and vocational paths, dignitary harms, and economic losses. Using an AI system to serve up targeted advertisements could cause a range of effects, from customer satisfaction to annoyance to manipulation to substantial financial impact and fraud. Using an AI system to screen future employees could result in efficient matching of employers to employees or in worsening racial and gender discrimination.

Each of these impacts might seem to sound in a different part of the legal system: health law, student privacy, consumer protection, employment law, tort liability, etc. And with different harms and different settings come different approaches to standing, to damages, to the professional and legal obligations of AI's human operators, and to what precautions are considered socially necessary. Thus it is clear that there will not be one single general "law of AI."

There is increasing convergence among regulators, however, over the idea that there is something about AI systems, just as there is something about data processing, that might merit a unified baseline approach. A class of concerns about AI systems resonates across use environments, including that such systems are problematically opaque, that their use can allow humans to evade accountability, that they can reify and amplify existing discrimination and disparities, that they can make illegitimate and unjustified decisions, that they can eliminate discretion and handle edge cases poorly, that their use can erode human expertise, and that they are prone to cascading failure. This is to say nothing of the fact that AI systems are inherently data processing systems, often raising questions about the provenance, use, and harms of data—questions that sound in data protection law.

Through a number of soft law instruments, a set of core principles have been emerging for what many call "trustworthy AI".[21] There are many similarities across these instruments, although they differ

---

[20] David Lehr & Paul Ohm, "Playing With the Data: What Legal Scholars Should Learn About Machine Learning," *U.C. Davis Law Review* 51 (2017): 653.

[21] See, e.g., Model AI Governance Framework, supra note 16, at 64 (compiling ethical frameworks); AI Risk Management Framework Second Draft, Aug. 18 2022, https://perma.cc/49PC-6UN8 (hereinafter NIST AI RMF 2nd Draft), Table 1, at 12 (referencing the OECD AI Recommendation, the EU AI Act, and EO 13960 as sources for principles of trustworthy AI); Carlos Ignacio Guttierez & Gary Marchant, "A Global

meaningfully as to the rights that should be afforded to an individual affected by an AI system. According to the NIST AI Risk Management Framework, trustworthy AI is "valid and reliable, safe, secure and resilient, accountable and transparent, explainable and interpretable, privacy enhanced, and fair with … harmful biases managed."[22] The opposites of these principles entail the harms that regulators are trying to prevent. That is, "untrustworthy AI" is unreliable, unsafe, unsecure, brittle, prone to cascading failure, opaque, unexplained, not interpretable, privacy-violating, unfair, and biased.

THE DEVELOPING LAW(S) OF AI

Proposals for regulating AI systems across fields have thus come to share core goals.[23] Regulators aim to eliminate or mitigate AI harms and typically write the new laws of AI to address the known problem of "garbage in, garbage out": that an AI system that is trained on junky data sets will reproduce said junkiness in its output. One of the most well-known examples of this is in facial recognition, where a team of AI researchers at the Massachusetts Institute of Technology found that facial recognition algorithms trained on white male faces were notably bad at correctly identifying black women.[24] Regulators typically also aim to root out and prevent discrimination emerging from other sources, whether deliberate discrimination cloaked in math or careless discrimination resulting from a programming decision. Regulators often attempt to restore accountability by requiring transparency of varying kinds,[25] requiring third-party oversight, or harnessing internal corporate and government accountability regimes. And regulators sometimes suggest or require that there should be quality testing and monitoring for failure once a given system is deployed.

In short, although there may never be a truly one-size-fits-all law of AI, regulators have been developing a toolkit for baseline regulation. Roughly speaking, that toolkit requires the developers of AI to pay attention to their data sets, conduct an impact assessment, mitigate harms, and maybe disclose some things to someone, whether that someone is a regulator or an auditor or an impacted person. That, in a nutshell, is the new law of AI.

Perspective of Soft Law Programs for the Governance of Artificial Intelligence," at 3 (finding 634 sources of soft law on AI governance in existence before 2019) (2021), https://perma.cc/TA8N-PJV5.

[22] NIST AI RMF 1.0, at 3.

[23] For the argument that regulators should be addressing well-established human rights instead of these "AI ethics," see Mantelero, supra note 13.

[24] Joy Buolamwini & Timnit Gebru, "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification," *Proceedings of Machine Learning Research* 81 (2018): 77.

[25] For an overview of different flavors and shapes of transparency in algorithmic accountability, see Margot E. Kaminski, "Understanding Transparency in Algorithmic Accountability," in *The Cambridge Handbook of the Law of Algorithms*, ed. Woodrow Barfield (Cambridge University Press, 2020).

This section expands the nutshell. It offers three examples of the developing law(s) of AI: Singapore's Model AI Governance Framework, the draft EU AI Act, and the NIST AI Risk Management Framework. Two of these examples are soft law (Singapore and NIST), and one has yet to be enacted (EU). But the convergence around not just risk regulation but a particular light-touch version of risk regulation is telling, as is the fact that more examples are readily available, in both the United States and the EU.[26] Sector-specific laws, too, have embraced this particular light-touch version of risk regulation as the dominant approach to governing AI.[27]

Why have regulators turned to risk regulation to govern AI? They have done so in part because of analogies drawn to other areas of the law that also use risk regulation.[28] The harms of AI systems, like the harms of data privacy and data security violations, structurally resemble, for example, environmental harms.[29] AI harms are often systemic and societal rather than solely individual. Tracing causality could present significant challenges to litigators for both environmental and AI harms. Like environmental harms, it can be hard to measure AI harms, yet AI systems can also cause mass damage to society. Injured parties, as in the environmental context, often face a collective action problem. In both contexts, some harms are really risks of future harm rather than vested injuries. For these reasons and more, a

[26] Andrew D. Selbst, "An Institutional View of Algorithmic Impact Assessments," *Harvard Journal of Law & Technology* 35 (2021): 117 (discussing Canada's approach to impact assessments); SR-11-7, see Andrew Burt, "Leave A.I. Alone," *New York Times*, Jan. 4, 2018 (discussing SR-117: "In the financial sector, for example, the Federal Reserve enforces a regulation called SR 11-7, which addresses the risks created by the complex algorithms used by today's banks. SR 11-7's solution to those challenges is called 'effective challenge,' which seeks to embed critical analysis into every stage of an algorithm's life cycle — from thoroughly examining the data used to train the algorithm to explicitly outlining the assumptions underlying the model, and more."); Jennifer D. Oliva, "Dosing Discrimination," *California Law Review* 110 (2022): 47; W. Nicholson Price II, "Regulating Black-Box Medicine," *Michigan Law Review* 116 (2017): 421; Gina-Gail S. Fletcher & Michelle M. Le, "The Future of AI Accountability in the Financial Markets, *Vanderbilt Journal of Entertainment & Technology Law* 24 (2022): 101.

[27] New York City passed a law requiring audits of automated decision-making in hiring. See J. Edward Moreno, "New York City AI Bias Law Charts New Territory for Employers," Bloomberg Law, Aug. 29, 2022, https://news.bloomberglaw.com/daily-labor-report/new-york-city-ai-bias-law-charts-new-territory-for-employers. Colorado recently passed a facial recognition law requiring "accountability reports" from government actors using facial recognition software. See https://perma.cc/Q6XC-EN2M.

[28] Kaminski, "Regulating the Risks of AI," supra note 8 (discussing risk regulation as legal transplant).

[29] A. Michael Froomkin, "Regulating Mass Surveillance as Privacy Pollution: Learning from Environmental Impact Statements," University of Illinois Law Review (2015): 1713, 1757–58; Omri Ben-Shahar, "Data Pollution," Journal of Legal Analysis 11 (2019): 104; Dennis Hirsch, "Protecting the Inner Environment: What Privacy Regulation Can Learn from Environmental Law," Georgia Law Review 41 (2006): 1–63.

number of scholars, advocates, and regulators have turned to environmental law as a possible model for AI law, inasmuch as (some) environmental law is ex ante, systemic, and regulatory in nature.

More specifically, they have turned to risk regulation, with a focus on impact assessments and risk mitigation.[30] Risk regulation can mean many things, including both a set of goals and a set of tools. While almost any law could be characterized as addressing or regulating risk, risk regulation as discussed here refers to a narrower category of regulation. Typically, risk regulation takes as its goal measuring, mitigating, and accepting risks in exchange for some sort of social benefit. Risk regulation as discussed here contrasts with, for example, law that aims to prevent any harms (that is, bans) and law that aims to compensate for individual harms (that is, liability regimes).

To be clear, well-designed risk regulation can—indeed, I argue it should—contain elements of both precautionary regulation and liability.[31] That is, the distinction between risk regulation qua risk regulation and precautionary approaches or liability is not so sharp as detractors of the latter two approaches might claim. Nevertheless, the AI risk regulation discussed here is largely ex ante, systemic, and concerned with society-wide (rather than individual) outcomes. It tends to favor risk analysis and mitigation, and to largely ignore both precautionary tactics, such as licensing or sandboxing, and postmarket measures. But risk regulation does not have to look this way. Other versions of risk regulation in other fields can look quite different. I return to this point below.

By far, the most common regulatory instrument in the new law of AI is the algorithmic impact assessment. A tool originating in environmental, data protection, and human rights regulation, an algorithmic impact assessment typically requires a company or government entity to identify, document, assess, and often mitigate risks before releasing a technology into the world. Some laws envision the impact assessment as a static, one-time exercise before a system is released. Others characterize it as inherently iterative and ongoing—a process rather than a document. Proponents argue that a good impact assessment process can result in significant risk-mitigation, better and more deliberate organizational values, public accountability (sometimes), and feedback for policymakers at a nascent stage of regulation (sometimes). Critics point out that impact assessments can in practice be a meaningless box-ticking exercise, empty corporate compliance that is little more than heavy navel-gazing.

---

[30] See, e.g., Dillon Reisman, Jason Schultz, Kate Crawford, & Meredith Whittaker, "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability (2018), https://perma.cc/J6YY-S733; Andrew D. Selbst, "Disparate Impact in Big Data Policing," *Georgia Law Review* 52 (2017): 109, 169.

[31] Douglas A. Kysar, "The Public Life of Private Law: Tort Law as a Risk Regulation Mechanism," *European Journal of Risk Regulation* 9 (March 2018): 48; Wendy E. Wagner, "When All Else Fails: Regulating Risky Products Through Tort Litigation," *Georgetown Law Journal* 95 (2007): 693. See also Boyd, supra note 9.

Drafters of the new law of AI often also include data quality measures: procedural and substantive requirements aimed at ensuring the use of high-quality data sets, either within an impact assessment process or alongside it. Regulators typically encourage or require some form of harm mitigation. Finally, the developing laws of AI tend to incorporate varying forms of accountability, from requiring an explanation of an AI decision to affected individuals, to requiring impact assessments to be made public or released to a regulator, to requiring input from impacted stakeholder groups, to requiring external audits.

Singapore's Model AI Governance Framework

Singapore's Model AI Governance Framework is one of the earlier models for AI risk regulation. The first draft of the Singapore framework was released at the World Economic Forum in Davos, Switzerland, in January 2019; the second draft was released one year later, also at Davos, in January 2020.

The Singapore framework is soft law: It is characterized as guidance.[32] But it is not entirely untethered from law on the books. Companies may use their adoption of the framework to demonstrate that they are in compliance with Singaporean data privacy law.[33]

The Singapore framework is based on two guiding principles: (a) that AI decision-making should be explainable, transparent, and fair; and (b) that AI solutions should be human centric, aimed at amplifying human capabilities and protecting the interests of human beings.[34] The Singapore framework provides guidance on four areas: internal governance, human involvement in AI-augmented decision-making, operations management, and stakeholder interaction and communication.[35]

Much of the Singapore framework's risk management approach to governance is developed in the first area: internal governance structures and measures. There, the framework suggests that risks associated with AI could be "managed within [an existing] enterprise risk management structure."[36] The framework encourages organizations to use risk management.[37] This includes assessing data sets and reviewing them for risks of inaccuracy or bias. It also includes establishing monitoring and reporting systems. As

[32] Model AI Governance Framework, supra note 16 at 7.

[33] Id. at 17, 2.12.

[34] Id. at 15.

[35] Id. at 20.

[36] Id. at 21.

[37] Id. at 24.

an additional aspect of risk management, the framework suggests designing AI systems to report the confidence level of their predictions.

The Singapore framework places a heavy emphasis on internal organizational governance and structures. It encourages organizations to use an "iterative and ongoing process … to continually identify and review risks relevant to their technology solutions, mitigate those risks, and maintain a response plan should mitigation fail."[38] Organizations should document this iterative process through an impact assessment that is reviewed periodically.[39]

The Singapore framework also addresses risk in its discussion of human involvement in AI-augmented decision-making. The framework characterizes decisions about human oversight as an aspect of its risk management approach.[40] The higher the risk—defined as the severity of harm multiplied by the probability of harm—the more organizations are encouraged to include human involvement in AI decision-making.

Elsewhere, the Singapore framework emphasizes data quality and the problems of biased data sets as factors that must be addressed to mitigate the risk of unintended discrimination.[41] The framework calls for "good data accountability practices," which include tracing where the data came from (data lineage) and maintaining a data provenance record; ensuring data quality; minimizing inherent bias in data sets, including both selection and measurement bias; using different data sets for training, testing, and validation; and periodically reviewing and updating data sets.[42]

The Draft EU AI Act

The draft EU AI Act is both different from and strikingly similar to Singapore's Model AI Governance Framework. Unlike the Singapore framework, the draft EU AI Act will be hard law.[43] That is, it will be enforced by centralized, top-down regulators and comes with significant potential penalties. In other ways, however, the EU AI Act takes a softer approach to governance, allowing governed companies to self-certify and encouraging them to participate in the creation of substantive standards. And at its core,

---

[38] Id. at 29.

[39] Id. at 29.

[40] Id. at 30.

[41] Id. at 36.

[42] Id. at 37–41.

[43] European Commission, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, at Title I, Art. 3(1), Annex I, COM (2021) 206 final (April 21, 2021).

like the Singapore framework, the EU AI Act uses risk regulation centered on risk assessments and mitigation to govern AI.

The EU AI Act primarily regulates the providers of AI systems. The act classifies the use of AI systems into three buckets: unacceptably risky, high-risk, and low- or minimal-risk uses of AI. The act bans certain uses of AI that raise unacceptable risks, including social scoring by public authorities,[44] law enforcement use of real-time biometrics in public spaces,[45] and uses of AI that subliminally distort a person's behavior in a manner that causes physical or psychological harm.[46] The EU AI Act subjects high-risk uses of AI to risk regulation. For the remaining low-risk uses of AI, the act encourages self-governance modeled on its risk regulation provisions.

The core of the AI Act consists of the risk regulation that governs high-risk AI systems. First, an AI provider must follow both the substantive and procedural requirements of the act. The act contains a number of substantive requirements, including for example requirements on data quality and accuracy, and delegates the content of other substantive requirements to technical standards-setting organizations, the European Commission, and even to implementing firms.

The AI Act also contains procedural requirements that largely constitute risk regulation. The AI Act requires providers to establish a "risk management system" that identifies risks, tests AI systems premarket, adopts "suitable risk management measures," and conducts postmarket monitoring. The act's procedural requirements also emphasize transparency and record-keeping. These requirements are aimed at increasing the quality of data sets and algorithms and increasing accountability. In addition to ex ante risk management, the AI Act requires premarket registration and postmarket monitoring and reporting, coupled with government oversight. Thus a key difference between the Singapore framework and the AI Act is government oversight and potential enforcement once the system is in use.

The EU AI Act also has its version of the algorithmic impact assessment. In addition to the above requirements, an AI provider under the act must undertake what is called a "conformity assessment" before releasing the AI system on the EU market. There are two distinct conformity assessment tracks for high-risk AI systems. If an AI system is part of a product that is already regulated for safety concerns, it undergoes the same conformity assessment process as other EU-regulated products. This entails going to a designated independent third party (a "notified body") for approval. If, however, the harms of the AI system are "mainly fundamental rights implications," then the AI provider essentially

[44] Id., Title III, Article 5(1)(c).

[45] Id., Title III, Article 5(1)(d).

[46] Id., Title III, Article 5 (1)(a).

self-certifies to the conformity assessment process outlined in the act.[47] That is, AI systems that implicate fundamental rights rather than safety are, strangely, regulated less strictly.[48]

NIST's AI Risk Management Framework

In 2020, Congress directed NIST to establish a risk management framework for AI systems.[49] If the EU AI Act exemplifies in some ways the European approach to governing AI, the NIST AI Risk Management Framework exemplifies a U.S. approach, which more closely resembles the Singapore framework.[50] Like the Singapore framework, the NIST approach is soft law, rather than command-and-control; it is intended for voluntary use, and there is no enforcement mechanism.[51] And like the Singapore framework, the NIST approach focuses primarily on enterprise risk management, rather than on establishing or protecting individual fundamental rights, or establishing regulatory infrastructure.[52] It is intended to be iterative in nature, to change over time.[53]

There are some similarities between the EU AI Act and NIST's AI Risk Management Framework, and also some significant differences. Once again, the core similarity is that both approaches focus on risk

[47] Id. at 13.

[48] For the reasoning behind this, see Id. at 14 ("As regards stand-alone high-risk AI systems that are referred to in Annex III, a new compliance and enforcement system will be established. This follows the model of … legislation implemented through internal control checks by the providers with the exception of remote biometric identification systems that would be subject to third party conformity assessment. A comprehensive ex-ante conformity assessment through internal checks, combined with a strong ex-post enforcement, could be an effective and reasonable solution for those systems, given the early phase of the regulatory intervention …. By contrast, for reasons of consistency with the existing product safety legislation, the conformity assessments of AI systems that are safety components of products will follow a system with third party conformity assessment procedures already established under the relevant sectoral product safety legislation."

[49] H. Rept. 116-455—Commerce, Justice, Science, and Related Agencies Appropriations Bill, 2021, and Section 5301 of the National Artificial Intelligence Initiative Act of 2020 (Pub. L. 116-283).

[50] NIST AI RMF 1.0.

[51] Id. at 2.

[52] Id. at 8 ("AI risk management should be integrated and incorporated into broader enterprise risk management strategies and processes.").

[53] Id. at 42 (The AI RMF is a "living document … [that] should be readily updated[.]").

management: on identifying, mitigating, and monitoring risks. Like the EU AI Act, NIST characterizes AI risk management as an iterative, ongoing process rather than a one-time checklist to be followed.[54]

The core difference between the U.S. and EU approaches is that the EU AI Act is centralized regulatory hard law, including bans, penalties, and ongoing government oversight. Rather than top-down formal law, NIST envisions AI risk management as an aspect of existing organizational risk management. While NIST's framework is likely to have an impact on industry practices, none of it is legally enforceable or required.

The NIST framework is also considerably shorter than the EU AI Act, with the "core" of the proposal clocking in at around ten pages, compared to the draft EU AI Act's seventy or so pages. The NIST framework adopts a similar core framework to past NIST guidance on, for example, cybersecurity.[55] The NIST framework consists of three cyclical functions: mapping, measuring, and managing. The framework dictates that organizations should first map their risks: find, recognize, and describe them.[56] Next, organizations should measure their risks, quantifying them where possible.[57] Finally, organizations should manage their risks, including by considering viable non-AI alternative systems, or by deciding to deactivate a system.[58] Throughout, the NIST framework emphasizes "a culture of risk management," with accountability structures that encourage challenges to risky designs and communications about risks.[59]

THE LIMITATIONS OF AI RISK REGULATION

I here offer three critiques that identify the limitations of AI risk regulation. First, risk regulation comes with known problems, which I refer to as "policy baggage," that are now being ported into the regulation of AI systems. Second, not all risk regulation is the same, and the version of risk regulation that dominates these new laws of AI runs the risk of not satisfying key stakeholders and not solving key problems. Indeed, the version of risk regulation repeatedly being deployed to regulate AI neglects a number of arguably essential tools in the risk regulation toolkit. Third, choosing to use risk regulation as

[54] Id. at 20 ("Risk management should be continuous, timely, and performed throughout the AI system lifecycle dimensions.").

[55] See Framework for Improving Critical Infrastructure Cybersecurity ("Cybersecurity Framework"), archived at https://perma.cc/7WQP-PAWC.

[56] NIST AI RMF 1.0, supra note 19 at 20, 25.

[57] Id. at 28.

[58] Id. at 32.

[59] Id. at 21.

an approach inaccurately assumes as a starting point that all AI systems can be fixed—that a more accurate/fair/transparent system would not itself cause social harms.

Policy Baggage

Risk regulation as applied in other fields has well-known pathologies, or policy baggage.[60] Risk regulation typically, though not always, is used to address measurable, quantifiable harms. This leads to three problems: (1) Harms that are harder, or impossible, to quantify are devalued; (2) unknown unknowns can get swept aside; and (3) normative values can be obscured within ostensibly objective "scientific" decisions. Using risk regulation in lieu of, rather than in conjunction with, tort liability also neglects some benefits that complimentary liability can offer.

Using risk regulation can deprioritize rights-based harms, such as those to dignity or autonomy, and harms that are otherwise harder to measure, such as emotional harms or harms to democratic society. Scholars have criticized impact assessments along these lines, pointing out that certain easier-to-measure harms are often valorized over others.[61] Take facial recognition, for example. Let's say a government agency wants to map, measure, and mitigate the harms of a facial recognition system. The agency would be more likely to mitigate harms that it can measure, such as a rate of error, than harms it cannot, such as the harms of pervasive public surveillance. Questions of morality and fairness—precisely the kinds of questions that governance of AI attempts to address—can be particularly challenging for risk regulation.

Research on risk regulation shows that regulators who deploy it often struggle to embrace epistemic humility. That is, regulators can struggle to admit that there are things they do not, and sometimes cannot, know. In the nuclear energy context, consider the example of a group of regulators who acted based on the assumption that solid nuclear waste, once buried in a repository, would not escape and harm the environment.[62] They acted as though the risk were zero, despite knowing that they did not know this for sure.[63] It turns out they were wrong: The rock at the containment site had fractures through which water could permeate, and plutonium is capable of traveling in water.[64] AI systems are rife with

---

[60] Kaminski, "Regulating the Risks of AI," supra note 8 at 21.

[61] Jacob Metcalf, Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, & Madeleine Clare Elish, "Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts," *Proceedings of the 2021 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2021).

[62] Daniel A. Farber, "Uncertainty," *Georgetown Law Journal* 99 (2011): 901, 910–11.

[63] Id. at 911 ("In short, because the probability of a release was not known and the agency felt optimistic, it decided simply to ignore the problem.").

[64] Id. at 950.

both known unknowns and unknown unknowns; they are complex systems, and complex systems can crash in unknown ways. If regulators lack epistemic humility, AI systems can and will surprise us, not always in good ways.

Risk regulation's quantitative leanings also mean that normative decisions can be subsumed into technical, technocratic conversations. Wendy Wagner has written about how seemingly neutral scientific decisions can mask significantly value-laden judgments, shielding those normative judgments from democratic accountability.[65] We see this policy baggage, too, playing out in the AI governance context, where normative judgments over what constitutes "fairness" are couched as mathematical debates, only later unearthed as value-laden and policy-based in nature.[66]

Risk regulation brings with it other policy baggage as well. Risk regulation typically is not aimed at providing either compensation for injured people or civil recourse in the form of individual process. Risk regulation's typical lack of compensation schemes has clear consequences, both as an inadequate deterrent for offenders and in failing to make injured people whole. Individual process can serve an important role beyond compensation, legitimizing a system and affording affected individuals dignity. For this reason, the newly proposed White House Blueprint for an AI Bill of Rights, discussed below, includes both notice and procedural rights to challenge AI systems' decisions and ask for a reconsideration, joining a number of countries or entities that call for an individual right to contest AI decisions.

Most proposals for AI risk regulation ignore tort law and focus on impact assessments. This is not to say they preempt tort law, but rather that tort law isn't being considered as a matter of regulatory design, and in some areas won't provide relief. As a matter of regulatory design, tort liability can serve as an important part of a feedback loop, in which the substantive output of civil cases can make its way into regulation.[67] Tort liability can also serve an essential information-forcing function.[68] Risk regulation risks being opaque and thus unaccountable; information is often hard for even regulators to obtain, let alone understand; and risk regulation by itself does not typically provide substantial avenues for nonexperts to obtain information or otherwise get involved. Tort liability, by contrast or in complement,

---

[65] Wendy E. Wagner, "The Science Charade in Toxic Risk Regulation," *Columbia Law Review* 95 (1995): 1613.

[66] See, e.g., Pauline T. Kim, "Auditing Algorithms for Discrimination," *University of Pennsylvania Law Review Online* 166 (2017): 189, 197–202; Deborah Hellman, "Measuring Algorithmic Fairness," *Virginia Law Review* 106 (2020): 811, 834 (outlining two conflicting ways to measure algorithmic fairness).

[67] Douglas A. Kysar, "The Public Life of Private Law: Tort Law as a Risk Regulation Mechanism," *European Journal of Risk Regulation* 9 (March 2018): 48.

[68] Wendy E. Wagner, "When All Else Fails: Regulating Risky Products Through Tort Litigation," *Georgetown Law Journal* 95 (2007): 693.

can serve an information-forcing function, getting "smoking-gun" information out into the public. Wagner has written about how class-action lawsuits have served an information-forcing function, pushing regulators to respond to public outrage after financially motivated plaintiffs' attorneys uncovered information that a regulatory regime missed. Douglas Kysar has written about a similar dynamic in environmental law. Without tort law, risk regulation can be static or, even worse, can be captured by regulated entities.

Inadequate Risk Regulation

In addition to the policy baggage that it brings, the use of risk regulation to govern AI systems has been leading to a now predictable set of conflicts. It turns out that not all risk regulation is the same. When policymakers in the United States think of risk regulation, they typically think of a highly quantitative version of cost-benefit analysis that weighs risks against benefits and regulates accordingly. When policymakers in the United Kingdom think of risk regulation, they typically think of a top-down model of regulation that identifies riskier actors or practices and allocates government resources by risk. Risk regulation in a particular jurisdiction or field can change over time, for example shifting from an emphasis on precaution to cost-benefit analysis, or vice versa.[69] Risk regulation can differ in different fields, depending on institutional and social histories, and variations in the types of harms, goals, and expertise at play. It differs in different countries. The point is: Just because legislators have decided to deploy risk regulation does not mean they will deploy the kind of risk regulation stakeholders are demanding.

The developing law of AI has been dominated by the enterprise risk management model of risk regulation. This version, like Singapore's Model AI Governance Framework and the NIST AI Risk Management Framework, leans heavily on internal governance and infrastructure to mitigate risks to businesses. This differs from what stakeholders appear to want: risk regulation modeled on environmental law, which entails public transparency and public participation. A now-predictable conflict has emerged over how much to delegate risk management to private companies (the enterprise risk management model) versus how much to deploy risk management as a form of democratic oversight (the environmental law, or NEPA [National Environmental Policy Act], model). At stake is a question of the central goal of regulation: Is it to mitigate risks to businesses (enterprise risk management), which means mitigating certain risks to impacted persons, or to create external oversight over systems' development and use (NEPA)?

This insight that not all risk regulation is the same leads to the observation of just how myopic and path-dependent AI risk regulation has already become. That is, while there are some variations in AI risk regulation proposals, they largely all follow a similar model: Watch your data sets, assess risks, do an

[69] Vogel, supra note 10.

impact assessment, do some risk mitigation—and don't establish or affirm a private right of action for related harms.

The risk regulation toolkit, deployed in areas of the law from toxic risk regulation to food and drug regulation, is much broader. AI risk regulation largely (though not entirely[70]) tends to ignore the following tools: bans, licensing, regulatory sandboxing, fail-safe modes, and postmarket measures, including conditional licensing, guardrail requirements, and postmarket monitoring. AI risk regulation also typically doesn't yet deploy substantive performance standards, largely kicking the can down the road.[71] The lack of clearly available tort liability for non-safety-related failures (read: privacy, discrimination, fairness, and even error with non-safety-related consequences) of AI systems, at least in the United States, means AI risk regulation largely lacks the backstop, feedback loop, and compensation schemes of tort law.

The Limits of Techno-Correctionism

AI risk regulation typically presumes that systems can be fixed. With the notable exception of the EU AI Act, most attempts at AI risk regulation do not seriously contemplate banning the use of any AI systems.[72]

The central attempt to "fix" AI systems evidences what Jessica Eaglin calls a "techno-correctionist" approach to regulating AI.[73] Techno-correctionism, per Eaglin, identifies the problems with AI systems and then uses regulation to try to fix often technical problems, ignoring bigger questions of what it means to design and use the technology in particular contexts in the first place. That is, techno-correctionism misses the fact that the design, aims, and uses of AI systems are political choices.

---

[70] The Draft EU AI Act contemplates bans, a sort of licensing-lite, and postmarket measures. Several other proposed U.S. laws contemplate either bans (Washington Senate Bill 5116) or substantive requirements (Washington Senate Bill 5116 and to some extent the proposed Algorithmic Accountability Act).

[71] Sandra Wachter & Brent Mittlestadt, "A Right to Reasonable Inferences: Re-thinking Data Protection Law in the Age of Big Data and AI," *Columbia Business Law Review* 2 (2019): 494, 572–88. But see Alicia Solow-Niederman, "Information Privacy and the Inference Economy," *Northwestern University Law Review* 117 (2022): 357, 420 ("[P]rocedural guidance only goes so far when it comes to AI-powered tools. Fairness in ML is hotly contested. There are no "accepted statistical principles and methodology" in many ML contexts; rather, the very choice of a mathematical definition of "fairness" is a political one[.]").

[72] And even the EU AI Act's bans have been criticized as narrow and static, creating cliff effects where one use of a system might be banned (e.g., facial recognition in real time in public places by law enforcement), while another closely related use might be permitted (e.g., facial recognition in public places by private actors), potentially leading to regulatory arbitrage.

[73] Eaglin, supra note 14.

For example, one problem with facial recognition is that it might be inaccurate or biased. Techno-correctionist regulation aims to reduce inaccuracy and "correct" bias. However, the use of accurate facial recognition to surveil and track people in public spaces is a political choice, with significant consequences for collective privacy, civil liberties, anti-discrimination values, and even democracy itself.[74] The use of accurate sentencing algorithms (assuming such a thing could be achieved), too, is political. It replaces other, more procedurally oriented and individualized approaches to sentencing, with consequences for theories of justice that emphasize individuality.[75] Techno-correctionism can thus obscure these political questions as to whether an AI system should be used at all. It can elide problems as to whether a system is being "fixed" toward a normatively problematic purpose.

The techno-correctionist approach often ignores, too, that the data used in AI decision-making is often social by nature, not some ground truth about the world. That is, the choices society has made and continues to make as to what crimes merit incarceration, where to concentrate policing, how much discretion to delegate to police and to prosecutors, and what resources are afforded defendants are all social elements that underpin the "fact" of a criminal conviction.[76] If a recidivism risk or predictive policing algorithm is trained on data that are constructed by past societal choices, those past societal choices will be perpetuated and reified. For example, if in the past, minor crack cocaine or marijuana offenses led to incarceration under a zero-tolerance policy, then an algorithm will get "good" at predicting who was likely, under this past policy, to get entrapped in the carceral system. Our past social facts thus risk becoming our future social facts, even if an algorithm is technically "accurate."

Underpinning the techno-correctionist assumption is a particular and often unquestioned epistemology of AI systems: the notion that there is some underlying ground truth that these computer programs can always uncover. This epistemology of AI presumes a certain neutrality to the nature of the technological instrument: It's "just math."[77] In some settings—for example, tracking and predicting baseball player

---

[74] Maria Badillo, "Judge Declares Buenos Aires' Fugitive Facial Recognition System Unconstitutional," The Future of Privacy Forum, Sept. 20, 2022, https://perma.cc/C8YQ-YEWQ.

[75] Jessica Eaglin, "Population-Based Sentencing," *Cornell Law Review* 106 (2021): 353, 357 ("[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.").

[76] See, e.g., Elizabeth E. Joh, "Feeding the Machine: Policing, Crime Data, and Algorithms," *William & Mary Bill of Rights Journal* 26 (2017): 287; Jessica Eaglin, "Constructing Recidivism Risk," *Emory Law Journal* 67 (2017): 59, 76.

[77] Ifeoma Ajunwa, "The Paradox of Automation as Anti-Bias Intervention," *Cardozo Law Review* 41 (2020): 1671, 1686; Cathy O'Neil, *Weapons of Math Destruction* (Crown, 2016): 21; Ngozi Okidegbe, "Discredited Data," *Cornell Law Review* 107 (2022): 2007, 2052 (describing the use of algorithms as entrenching "epistemic oppression").

stats—accuracy grounded in measurable fact may be a realistic goal.[78] In others, however, where an AI system is used to model and predict a contested social concept on which there is much normative disagreement, "accuracy" is itself a loaded goal. It presumes that an AI system is predicting a brute fact rather than a social fact.[79]

POLICY SUGGESTIONS

The developing law of AI is predominantly risk regulation. There are clear benefits to trying to address AI systems' harms on a systemic level, before such systems are deployed.[80] But AI risk regulation as currently envisioned neither addresses the misfit between its nonquantitative goals and its pseudo-quantitative approaches, nor acknowledges the limitations of taking a techno-correctionist approach. This paper, probably unsatisfyingly, does not propose a single solution. Instead, it identifies room for growth and change. Regulators drafting omnibus AI laws should broaden their consideration of the problems posed by AI systems and, more pointedly, look both to other available tools in the risk regulation toolkit and beyond.

First and foremost, AI risk regulation needs to adopt some epistemic humility. There are some problems it cannot solve, some systems that should not be used at all. Some AI systems are snake oil.[81] For example, hiring algorithms that purport to analyze an applicant's "affect" or personality, what Ifeoma Ajunwa has deemed modern-day phrenology,[82] are not grounded in science and are typically discriminatory in nature.[83] Snake-oil AI systems should be banned, at least where they have significant effects on people. Or regulators might aim to put in place substantive accuracy requirements, such that system developers and users should have to show that they are using accepted methodology and building

[78] O'Neil, Id. at 17.

[79] Joh, supra note 76, at 295.

[80] Ifeoma Ajunwa, "Automated Video Interviewing as the New Phrenology," *Berkeley Technology Law Journal* 36 (2022): 101; Margot E. Kaminski, "Binary Governance: Lessons From the GDPR's Approach to Algorithmic Accountability," *Southern California Law Review* 92 (2019): 1529; Selbst, supra note 26, at 140.

[81] Inioluwa Deborah Raji, Aaron Horowitz, I. Elizabeth Kumar, & Andrew D. Selbst, "The Fallacy of AI Functionality," *Proceedings of the 2022 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2022). See also Arvind Naranyan, "How to Recognize AI Snake Oil," https://perma.cc/68JB-QYPZ.

[82] Ajunwa, supra note 80.

[83] See also Luke Stark and Jevan Hutson, "Physiognomic Artificial Intelligence," *Fordham Intellectual Property, Media and Entertainment Law Journal* 32 (2022): 922.

on sound research.[84] Such an accuracy requirement might constitute a ban of some systems in practice. There may be room here, too, for regulation of deceptive uses of AI systems, including through consumer protection regulators such as the Federal Trade Commission.[85]

There are other problems with the use of AI systems that fall through the cracks of the current risk regulation approach. Some uses of AI systems—such as the use of facial recognition to conduct continuous tracking through the surveillance of public spaces—implicate constitutional rights, or at least human rights values, that risk regulation largely does not consider.[86] Sometimes, AI risk regulation fails as problems arise when a system that was designed and tested under one set of conditions then gets used by undertrained users in another environment (for example, when judges rely on recidivism risk algorithms, or police officers rely on facial recognition, without understanding the potential for inaccuracy) or for different purposes.[87] And then there are uses of AI systems where all the well-meaning risk regulation in the world can't change the fact that the locus and process of decision-making has been shifted from a human individual, often trained with situational expertise, to ex ante human programmers in a different organizational environment and often lacking such site-specific expertise, making system-level decisions that can be both policy-laden and unqueried.[88]

AI risk regulation should not take as its starting point that light-touch risk regulation solves all. Regulators should not assume that add-on sectoral laws will take care of these problems and thus fail to address them in baseline regulation. The first suggestion of this paper is simple: Risk regulation should address its own limitations, acknowledging where it will not be enough.

The second suggestion of this paper is simple, too: Risk regulation in general has more tools at its disposal, and regulators should consider using them. One idea, apart from or in addition to bans, would

---

[84] See, generally, infra note 71.

[85] Andrew D. Selbst & Solon Borocas, "Unfair Artificial Intelligence: How FTC Intervention Can Overcome the Limitations of Discrimination Law," *University of Pennsylvania Law Review* 171 (forthcoming 2023); Woodrow Hartzog, "Unfair and Deceptive Robots," *Maryland Law Review* 74 (2015): 785.

[86] See, e.g., Andrew Guthrie Ferguson, "Facial Recognition and the Fourth Amendment," *Minnesota Law Review* 105 (2020): 101.

[87] Andrew D Selbst, danah boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, & Janet Vertesi, "Fairness and Abstraction in Sociotechnical Systems," *Proceedings of the 2019 Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2019): 59. To be fair, the NIST AI RMF contemplates this issue. See, e.g., NIST AI RMF 1.0 at 25.

[88] See, e.g., Danielle Keats Citron, "Technological Due Process," *Washington University Law Review* 85 (2008): 1249; but see Andrew Keane Woods, "Robophobia," *University of Colorado Law Review* 93 (2022): 51.

be to lean more heavily on licensing of various shapes and kinds, including conditional licensing. Andrew Tutt proposed, years ago, "an FDA for algorithms" that would approve algorithms for widespread distribution.[89] Gianclaudio Malgieri and Frank Pasquale have suggested more recently a modified licensing regime, wherein AI systems designers would pass certain hurdles and either seek regulatory approval or self-certify, with accountability, before they could release their systems for use.[90] They call this "unlawfulness by default" for certain AI systems, where AI developers must affirmatively demonstrate that their technology is not harmful before they deploy it. As Malgieri and Pasquale's proposal suggests, short of formal top-down licensing, there are other licensing-like tools that could be used to increase accountability. AI risk regulation could, for example, shift from pure self-assessment to third-party oversight for ex ante assessment and mitigation, as the EU AI Act does with respect to systems that pose risks to physical safety—but not for systems raising risks to human rights.

Conditional licensing, too, may have its place. Conditional licensing could involve licensing (or self-certifying) AI systems for use under only certain circumstances, or licensing (or self-certifying) AI systems for use only with built-in guardrails. In the context of AI systems, this could mean conditioning use of a system on adequate user training or restricting the use of a system designed for one purpose from use for another purpose. The idea of guardrails arises in several versions of AI risk regulation but typically is voluntary rather than required.

Regulators could use revocable licensing: offering licensing that gets withdrawn if and when harms surface. That is, short of taking a purely precautionary approach to AI systems, regulators could maintain a mechanism for recalling systems that are shown to be harmful or don't work. For revocable licensing to be effective, there must be some form of postmarket monitoring, as contemplated in the EU AI Act. Regulators can conduct postmarket monitoring or can impose mandatory reporting requirements on regulated entities. If regulators choose to rely on self-reporting for postmarket monitoring, there will need to be a spot-checking system, whether by regulators or third parties.

Regulators can use a variety of other tools not yet deployed across much of AI risk regulation. They could permit research into certain uses of AI systems, coupling such permissiveness with restrictions on

[89] Andrew Tutt, "An FDA for Algorithms," *Administrative Law Review* 69 (2017): 83.

[90] Gianclaudio Malgieri & Frank Pasquale, "From Transparency to Justification: Toward Ex Ante Accountability for AI," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4099657.

widespread public use.[91] They could use a regulatory sandboxing approach, allowing experimentation with the technology in controlled circumstances while also experimenting with regulatory approaches.[92]

Regulators could do better even within the existing risk-analysis-and-mitigation framework. Even if regulators choose not to implement formal licensing, or to shore up accountability for the licensing-lite regimes being proposed, they could at least put more of a substantive thumb on the risk assessment scale. Regulators could require companies to explicitly identify worst-case scenarios and weigh the best-case and worst-case outcomes.[93] They could require stress testing using worst-case scenarios or known historic data.[94] Regulators could require the use of scenario analysis and, in the face of uncertainty, require regulated entities to imagine multiple worst-case scenarios and plan for the worst of them.[95] Many of these are known tools of resilience regulation: regulation that accepts that there will be harms but aims at mitigating harms as they occur.[96]

As discussed at length above, a core concern with AI risk regulation is that it aims to reduce risks that cannot readily be quantified: risks of discrimination and "unfairness" and to privacy and other human rights and civil liberties.[97] Consequently, some harms get ignored, while others get devalued. One way to try to address this is to involve more, and more diverse, stakeholders in the assessment of AI harms.[98] Or we could arrive at substantive standards via legislation or public rulemaking, rather than delegating the interpretation of what constitutes harm to regulated entities.

---

[91] Farber, "Uncertainty," supra note 62, at 948 (discussing "restrictions on uses involving potential public exposure until further risk information is available, and sensitivity to potential large downside risks.").

[92] Sofia Ranchordas, "Experimental Regulations for AI: Sandboxes for Morals and Mores, in Morals and Machines," *University of Groningen Faculty of Law Research Paper* No. 7/2021 (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3839744.

[93] Farber, "Uncertainty," supra note 62, at 953, calls it "ambiguity theory," at 958.

[94] Id. at 957.

[95] Id. at 934–35. The goal of such scenario analysis is to eliminate strategies that violate minimum safe standards, and "locate strategies that function well under adverse circumstances." Id. at 935.

[96] Gary E. Marchant & Yvonne A. Stevens, "Resilience: A New Tool in the Risk Governance Toolbox for Emerging Technologies," *U.C. Davis Law Review* 51 (2017): 233.

[97] The regulation of AI systems is not the only place where risk regulation is applied to human rights; human rights impact assessments attempt to do much the same thing. Human rights impact assessments face similar issues to those discussed here. Mantelero, supra note 13.

[98] See Metcalf et al., supra note 61; Eaglin, supra note 76; Ngozi Okidegbe, "Discredited Data," *Cornell Law Review* 107 (2022): 2007.

Many areas of the law, including environmental law, use performance-based regulation.[99] Regulators articulate a standard, and regulated entities come up with ways to meet it. This is for the most part missing from AI risk regulation, as regulators have largely elected not to define what they mean by their terms. However, it could be done through substantive regulation. In Washington state, for example, a proposed law requires the meaning of "systemic discrimination" to be determined through regulations promulgated under the state administrative procedure act and in consultation with impacted stakeholder groups.[100] This kind of process would shift the struggles over assessing AI harms to a public conversation. Enterprise risk assessment and mitigation that occurs only within a firm or government entity can be simultaneously analytically unsatisfactory and democratically unsatisfactory, not to mention captured.[101]

AI risk regulation typically eschews both private rights of action and individual rights of other kinds. First, AI risk regulation does not provide compensation for harmed individuals. Risk regulation could include compensation schemes, even if it avoids tort law.[102] But tort can serve an important, even essential set of roles. I discussed the role of tort liability in risk regulation above: It can serve as a feedback loop, and it can serve as a regulatory backstop. With AI risk regulation, tort liability is probably largely lacking. In some areas of law, for example in safety-critical contexts, tort liability likely exists. With other kinds of harms, however, the availability of a private right of action through existing law might be limited by substantive doctrine or law[103] and by the Supreme Court's recent jurisprudence

---

[99] Lauren E. Willis, "Performance-Based Consumer Law," *University of Chicago Law Review* 82 (2015): 1309.

[100] The bill tasks the director of the state's Chief Information Officer with adopting rules that, among other things, will define "systemic discrimination." Section 3.

[101] Doug Kysar, "It Might Have Been: Risk, Precaution, and Opportunity Cost," *Journal of Land Use* 22 (fall 2006). Sometimes quantification and cost-benefit analysis (CBA) is "both analytically and democratically unsatisfactory". Id. at 8 (Touting the "underappreciated benefits to the [precautionary principle's (PP's)] more modest approach. … [U]nlike the optimization framework of CBA, which proceeds awkwardly in the absence of fully characterized risks and consensus normative agreement on exogenized choice criteria, the PP's approach reflects great sensitivity to the fact that decisionmaking in the face of many … problems demands not only substantive, but also *procedural* and *discursive* rationality.").

[102] See, e.g., Adam Thierer, "When the Trial Lawyers Come for the Robot Cars," *Slate,* June 2016 (suggesting a compensation scheme modeled on the National Childhood Vaccine Injury Act).

[103] Ifeoma Ajunwa, "The Paradox of Automation as Anti-Bias Intervention," *Cardozo Law Review* 41 (2020): 1671, 1726–27; Solon Barocas & Andrew D. Selbst, "Big Data's Disparate Impact," *California Law Review* 104 (2016): 671; Danielle K. Citron, "Mainstreaming Privacy Torts," *California Law Review* 98 (2010): 1805.

on standing.[104] These gaps in underlying liability have consequences, discussed above, ranging from the lack of a substantive feedback loop for regulation to a problem of obtaining adequate information flows from regulated entities.

AI risk regulation, too, largely ignores individual rights. A counter-strain articulated in AI soft law emphasizes in parallel to AI risk regulation a set of individual rights for those affected by AI systems. For example, the European Commission's High-Level Expert Group on Artificial Intelligence's 2019 report emphasized the centrality of fundamental human rights to AI regulation—what the group calls a "human-centric" approach.[105] This includes a right to explanation and the ability to contest decisions made using AI systems.[106] The Council of Europe, too, adopted a recommendation based on research by its own committee of experts that includes individual rights to transparency, contestability, and effective remedies.[107]

In the United States, the recently released White House Blueprint for an AI Bill of Rights echoes these documents, calling for notice and explainability and "access to timely human consideration and remedy by a fallback and escalation process if an automated system fails, it produces an error, or you would like to appeal or contest its impacts on you."[108] I have argued for these individual rights, not just as a way of addressing individualized harms that are not covered in systemic regulation, but as a necessary aspect of the governance of AI systems—a check on both regulators and regulated entities, and another source of both accountability and transparency if public transparency is lacking.[109]

The people and groups affected by AI systems are often missing from AI risk regulation. They lack avenues for redress, they aren't afforded process, and they have little say in what harms matter or count. As the European Data Protection Board and European Data Protection Supervisor stated in their critique of the draft EU AI Act: "Whether they are end-users, simply data subjects or other persons concerned by

---

[104] See, e.g., *TransUnion LLC v. Ramirez*, 141 S. Ct. 2190 (2021).

[105] High-Level Expert Group on Artificial Intelligence Set Up by the European Commission, "Ethics Guidelines for Trustworthy AI" (2019) at 12.

[106] Id. at 15 (discussing "the ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them").

[107] Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to Member States on the Human Rights Impacts of Algorithmic Systems, April 2020, archived at https://perma.cc/SL79-8N55.

[108] White House Office of Science and Technology Policy, "Blueprint for an AI Bill of Rights," October 2022, https://perma.cc/79QH-D46B.

[109] Kaminski, "Binary Governance," supra note 80.

the AI system, the absence of any reference in the text to the individual affected by the AI system appears as a blind spot."[110]

Indeed.

CONCLUSION

We are in the origin story of the law of AI. Through soft law and hard law instruments, there has been considerable consensus-building around the use of risk regulation to govern AI systems. There is a real temptation to see all of this as a first step on the way to more substantive regulation.[111] But there is also a strong risk of path dependency: that what we have now is all we are going to get.

Deploying risk regulation, whatever its benefits, is a normative choice with consequences. The type of risk regulation being used to regulate AI systems is light-touch law, centering on impact assessments and internal risk mitigation, and largely eschewing tort liability and individual rights, along with postmarket measures, licensing schemes, and substantive standards. It may well be worth looking at substantive areas of law that take AI risks more seriously—health law, or the law of financial systems—as potential models for better general AI risk regulation.

Multiple aspects of AI systems, and their harms, get lost in the current framing. AI is often the product of surveillance; this fact typically gets lost in discussions of AI risk regulation. The individual gets lost. The harms of AI systems can echo and reify problematic aspects of society. We should be asking, in the first place, whether we really want to measure, scale, and reproduce what are often deeply troubling aspects of existing social systems. The use of AI systems can leave less room for change, discretion, or compassion. There are big normative arguments to be had—and being had—on what this means for marginalized people and communities. These are policy conversations, not decisions to be left for enterprise risk management.

Regulating the risks of AI often means trying to "fix" AI. We should instead be asking bigger questions. Why is AI an increasingly go-to instrument of the carceral state?[112] Why is AI less likely to be deployed, for example, to translate court proceedings for asylum applicants?[113] Sometimes, the more important

---

[110] EDPB-EDPS 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) at 8 (June 2021), https://perma.cc/6AUM-NVLX.

[111] Selbst, "An Institutional View of Algorithmic Impact Assessments," supra note 26.

[112] Eaglin, "When Critical Race Theory Enters the Law & Technology Frame," supra note 14.

[113] Ryan Calo & Danielle K. Citron, "The Automated Administrative State: A Crisis of Legitimacy," *Emory Law Journal* 70 (2021): 797.

problem is not how to make the technology better, but what it means that we are using AI systems toward particular ends in the first place.

The Digital Social Contract paper series is supported by funding from the John S. and James L. Knight Foundation and Meta, which played no role in the selection of the specific topics or authors and which played no editorial role in the individual papers.