Fordham Law Review

Volume 88 | Issue 2 Article 3

2019

Liability for AI Decision-Making: Some Legal and Ethical Considerations

Iria Giuffrida
William & Mary Law School

Recommended Citation
Iria Giuffrida, *Liability for AI Decision-Making: Some Legal and Ethical Considerations*, 88 Fordham L. Rev. 439 (2019).
Available at: https://ir.lawnet.fordham.edu/flr/vol88/iss2/3

LIABILITY FOR AI DECISION-MAKING: SOME LEGAL AND ETHICAL CONSIDERATIONS

*Dr. Iria Giuffrida**

INTRODUCTION

Benjamin Franklin wrote that nothing is certain "except death and taxes."[1] A cynical former litigator, like the author, might add to those the certainty of litigation as new technology creates an increasing number of real challenges. With breakthroughs in artificial intelligence (AI) and related technologies, their uses are being implemented in government,[2] finance,[3] health care,[4]

* Visiting Assistant Professor of Law and Deputy Director, Center for Legal & Court Technology, William & Mary Law School. This Article is based on the author's presentation at the *Rise of the Machines: Artificial Intelligence, Robotics, and the Reprogramming of Law* Symposium on February 15, 2019, organized by the *Fordham Law Review* and Fordham's Neuroscience and Law Center. The author is grateful to Professor Fredric Lederer for his ongoing support and guidance, to Taylor Treece for her excellent research, and to Kazia Nowacki, Lindsey Whitlow, and James Lomonosoff for invaluable additional research. All errors are the author's own. This work was supported by a grant from the Silicon Valley Community Foundation, funded by Cisco Systems, Inc. For an overview of the Symposium, see Deborah W. Denno & Ryan Surujnath, *Foreword: Rise of the Machines: Artificial Intelligence, Robotics, and the Reprogramming of Law*, 88 FORDHAM L. REV. 381 (2019).

[1] Letter from Benjamin Franklin to Jean Baptiste Le Roy (Nov. 13, 1789), *in* 10 THE WRITINGS OF BENJAMIN FRANKLIN, 1789–1790, at 68, 69 (Albert Henry Smyth ed., 1907).

[2] There is a rich literature on the advancement of "smart cities." *See, e.g.*, Robert Brauneis & Ellen P. Goodman, *Algorithmic Transparency for the Smart City*, 20 YALE J.L. & TECH. 103, 114–15 (2018); Rob Kitchin, *The Real-Time City?: Big Data and Smart Urbanism*, 79 GEOJOURNAL 1, 1 (2014); Sofia Ranchordas & Abram Klop, *Data-Driven Regulation and Governance in Smart Cities*, *in* RESEARCH HANDBOOK IN DATA SCIENCE AND LAW 245, 245–73 (Vanessa Mak et al. eds., 2018).

[3] Megan Ji, Note, *Are Robots Good Fiduciaries?: Regulating Robo-Advisors Under the Investment Advisers Act of 1940*, 117 COLUM. L. REV. 1543, 1559 (2017); Tara Siegel Bernard, *The Pros and Cons of Using a Robot as an Investment Adviser*, N.Y. TIMES (Apr. 29, 2016), http://www.nytimes.com/2016/04/30/your-money/the-pros-and-cons-of-using-a-robot-as-an-investment-adviser.html [https://perma.cc/3ZQL-XW2Y].

[4] Bernard Marr, *How Is AI Used in Healthcare—5 Powerful Real-World Examples That Show the Latest Advances*, FORBES (July 27, 2018, 12:41 AM), https://www.forbes.com/sites/bernardmarr/2018/07/27/how-is-ai-used-in-healthcare-5-powerful-real-world-examples-that-show-the-latest-advances [https://perma.cc/L94B-WJRT]; Alvin Powell, *The Algorithm Will See You Now*, HARV. GAZETTE (Feb. 28, 2019), https://news.harvard.edu/gazette/story/2019/02/in-health-care-ai-offers-promise-and-hype/ [https://perma.cc/FF5H-D6DX]; *see also* Iria Giuffrida & Taylor Treece, *Keeping AI Under Observation: Anticipated Impacts on Physicians' Standard of Care*, 22 TUL. J. TECH. & INTELL. PROP. (forthcoming Fall 2019).

law,[5] environmental protection,[6] and education.[7]

AI plays varied functions in these applications. AI systems can be *descriptive* as they tell you what happened; *diagnostic* as they tell you why something happened; *predictive* as they forecast what will (statistically) happen; and *prescriptive* in being capable of performing actual decision-making and implementation.[8]

The creation and commercialization of these systems raise the question of how liability risks will play out in real life. However, as technical advancements have outpaced legal actions, it is unclear how the law will treat AI systems. This Article briefly addresses the legal ramifications and liability risks associated with reliance on—or delegation to—AI systems, and it sketches a framework suggesting how we can address the question of whether AI merits a new approach to deal with the liability challenges it raises when humans remain "in" or "on" the loop. This Article also suggests that questions of how we, as a society, deal with those challenges have to be connected to the broader ethical questions that AI evokes, such as whether we really want to create a fully autonomous system that we cannot control and how we could protect against "artificial stupidity."[9]

[5] For general legal practice, see, for example, Ed Walters, *AI Practice, Not Promise, in Law Firms*, A.B.A. (Jan. 1, 2019), https://www.americanbar.org/groups/law_practice/publications/law_practice_magazine/2019/january-february/JF2019Walters/ [https://perma.cc/NYR4-UTG8]. There is also a large offering of legal AI and analytics tools, such as Lexis Advance, Context, Lex Machina, and Westlaw Edge. In the criminal justice system, predictive systems—increasingly common tools designed to forecast recidivism rates—are causing ethical and practical challenges. *See generally* CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 3 U. BOLOGNA L. REV. 180 (2018); Han-Wei Liu et al., *Beyond* State v. Loomis*: Artificial Intelligence, Government Algorithmization and Accountability*, 27 INT'L J.L. & INFO. TECH. 122 (2019).

[6] *See generally* WORLD ECON. FORUM, HARNESSING ARTIFICIAL INTELLIGENCE FOR THE EARTH (2018), http://www3.weforum.org/docs/Harnessing_Artificial_Intelligence_for_the_Earth_report_2018.pdf [https://perma.cc/NCA8-YRLD]. The University of Southern California's Center for Artificial Intelligence in Society hosted on February 8, 2019, a symposium on AI for environmental conservation sponsored by the Microsoft AI for Earth program. For further detail on the symposium's agenda, see *Symposium on AI for Conservation*, USC CTR. FOR ARTIFICIAL INTELLIGENCE SOC'Y, https://www.cais.usc.edu/events/symposium-on-ai-for-conservation [https://perma.cc/AK9M-G264] (last visited Oct. 6, 2019).

[7] Bernard Marr, *How Is AI Used in Education—Real World Examples of Today and a Peek into the Future*, FORBES (July 25, 2018, 12:26 AM), https://www.forbes.com/sites/bernardmarr/2018/07/25/how-is-ai-used-in-education-real-world-examples-of-today-and-a-peek-into-the-future [https://perma.cc/J37N-DL2Z]; *see also AIED 2019: The 20th International Conference on Artificial Intelligence in Education*, AIED 2019, https://caed-lab.com/aied2019 [https://perma.cc/HZU7-JMGX] (last visited Oct. 6, 2019).

[8] Humberto Farias, *Machine Learning Vs Predictive Analytics: What's the Difference?*, CONCEPTA (Oct. 10, 2017), https://conceptainc.com/blog/machine-learning-vs-predictive-analytics-what-is-the-difference/ [https://perma.cc/6QAS-MB47].

[9] Nick Bostrom & Eliezer Yudkowsky, *The Ethics of Artificial Intelligence*, *in* THE CAMBRIDGE HANDBOOK OF ARTIFICIAL INTELLIGENCE 316, 316–34 (Keith Frankish & William M. Ramsey eds., 2014).

I. FACTORS IMPACTING LIABILITY RISK: AI SYSTEMS IN CONTEXT

Although the concept of AI has been around since at least the 1950s, it has only recently reached the social consciousness (other than through science fiction). Yet, no generally accepted definition exists.[10] In its most basic sense, AI refers to "the ability of a machine to perform cognitive functions we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem solving, and even exercising creativity."[11] The growth of AI has been fostered by technical advances in machine learning, which, rather unsophisticatedly, refers to the ability of an AI system to modify itself by taking into account new data.[12] Put differently, without explicit programming, the machine "statistically learn[s]"[13] from the data it has processed. This is what computer scientists call a shift from "deterministic to probabilistic computing."[14]

AI can be extraordinarily capable of analyzing enormous amounts of data to identify patterns and utilize them in a predictive or prescriptive sense.[15] For instance, when exposed to a sufficient number of cat and dog images, an AI system should then be capable of distinguishing between cat and dog photos. In addition, given enough accurate initial data, an AI facial recognition system should be able to identify a picture of an unknown person with impressive accuracy;[16] or, an AI may be able to predict future criminality.[17]

Although oversimplistic and certainly not adequately descriptive of the technical and sophisticated work necessary to create a useful and reliable AI system, the basic steps in the creation of an AI based on machine learning

[10] Calo, *supra* note 5, at 184–86; John McCarthy, *What Is Artificial Intelligence?*, STAN. U. 2–3 (Nov. 12, 2007), http://jmc.stanford.edu/articles/whatisai/whatisai.pdf [https://perma.cc/LRM5-UFKY].

[11] *An Executive's Guide to AI*, MCKINSEY & CO., https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/an-executives-guide-to-ai [https://perma.cc/6372-KDCM] (last visited Oct. 6, 2019).

[12] *Id.*

[13] Aditya Mohta, *The Role of Artificial Intelligence in Digital Transformation*, CISCO LIVE 4, https://www.ciscolive.com/c/dam/r/ciscolive/emea/docs/2019/pdf/PSODGT-2370.pdf [https://perma.cc/S865-ZWMD] (last visited Oct. 6, 2019).

[14] Noah D. Goodman, *The Principles and Practice of Probabilistic Programming*, 40 PROC. ANN. ACM SIGPLAN-SIGACT SYMP. ON PRINCIPLES PROGRAMMING LANGUAGES 399, 399–401 (2013).

[15] *See supra* note 8 and accompanying text.

[16] *See generally* Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACHINE LEARNING RES. 77 (2018). However, contemporary American efforts appear to often fail due to inadequate or biased training data. *See, e.g.*, *id.* at 77. The Chinese government's "social credit system" uses vast amounts of data and is intended to identify individuals, if necessary, by video image alone. *See, e.g.*, Vicky Xiuzhong Xu & Bang Xiao, *China's Social Credit System Seeks to Assign Citizens Scores, Engineer Social Behaviour*, ABC: NEWS (Apr. 1, 2018, 11:38 PM), http://www.abc.net.au/news/2018-03-31/chinas-social-credit-system-punishes-untrustworthy-citizens/9596204 [https://perma.cc/UWF3-SHC6].

[17] Julia Dressel & Hany Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, 4 SCI. ADVANCES, Jan. 2018, at 1, 1 (concluding that "widely used commercial risk assessment software COMPAS is no more accurate or fair than predictions made by people with little or no criminal justice expertise").

are: (1) coding of the underlying AI program; (2) training the AI to accomplish its function; and (3) ongoing self-modification by the AI based on changes in the underlying data and feedback loops.

The risk of AI error is huge.[18] Even if the basic program is accurate, training the program requires an adequate amount of correct and nonbiased data, a requirement that in our real world appears to be difficult to meet.[19] Once applied to real-world instances, an AI exposed to a sufficient amount of biased, false, or otherwise corrupted data may modify its "understanding" to accept the biased or false information as accurate and perform its function based on that erroneous data.[20] But there's more.

A. The "AI Ecosystem"

AI systems do not perform in an informational vacuum. AI is only one part of the multiple modern technologies which interact with each other and with the kinetic world around them. The key is the *combination* of all of these technologies and more that constitutes the "AI Ecosystem": AI systems interacting in a data-rich environment, fueled by the "internet of things," enabled by sensors, blockchain, and other technologies.[21] Data is constantly created, exchanged, analyzed, pooled, and reassessed.[22] Each of these technologies carries independent liability risks. When they combine (as they inevitably do in "real" life), the liability landscape becomes layered and increasingly complex.[23] In addition to these machine-to-machine relationships, embedded within that AI Ecosystem is the relationship between AI and humans, as data, programming, and usage can be affected by fallible human beings.

These features make AI and related technologies sublimely useful but also intrinsically problematic. Not only is it inherently difficult to determine why an AI system reached a given output or decision[24] but, because of how the AI Ecosystem operates, it may be impossible to reverse engineer the decision-making process to know on which data the AI system relied. This

[18] Chris Hoffman, *The Problem with AI: Machines Are Learning Things, but Can't Understand Them*, HOW-TO GEEK (Jan. 6, 2019, 10:03 PM), https://www.howtogeek.com/394546/the-problem-with-ai-machines-are-learning-things-but-cant-understand-them/ [https://perma.cc/5PFK-ALPW]; *cf.* Buolamwini & Gebru, *supra* note 16, at 77.

[19] Hoffman, *supra* note 18.

[20] Thomas C. Redman, *If Your Data Is Bad, Your Machine Learning Tools Are Useless*, HARV. BUS. REV. (Apr. 2, 2018), https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless [https://perma.cc/Q34U-GWLE].

[21] *See generally* Iria Giuffrida, Fred Lederer & Nicolas Vermeys, *A Legal Perspective on the Trials and Tribulations of AI: How Artificial Intelligence, the Internet of Things, Smart Contracts, and Other Technologies Will Affect the Law*, 68 CASE W. RES. L. REV. 747 (2018).

[22] *See generally id.*

[23] *See generally id.*

[24] Peter Holley, *How Quickly Can AI Solve a Rubik's Cube?: In Less Time Than It Took You to Read This Headline.*, WASH. POST (July 16, 2019), https://www.washingtonpost.com/technology/2019/07/16/how-quickly-can-ai-solve-rubiks-cube-less-time-than-it-took-you-read-this-headline [https://perma.cc/YF6S-J23A] ("But since the algorithm was programmed merely to solve the puzzle, researchers were left with a limited understanding of *how* it did so." (emphasis added)).

is the classic "black box problem" that reflects the lack of transparency and explainability that may render AI decision-making processes impenetrable.[25]

B. Classic Tort Law

In the United States, tort law has two principal normative aims: compensating tort victims and deterring future tortious conduct.[26] The focus of tort law is to determine who is liable for the loss suffered by the plaintiff as caused by the tortfeasor's wrongful act—an act from which it was reasonably foreseeable that losses would flow.[27] Tort suits involving harm caused by devices usually allege either negligence from the tortfeasor or are based on a theory of products liability.[28] Liability for a defective product applies when, among other possibilities, a reasonable alternative design could have prevented or limited foreseeable risks of harm.[29]

In addition, many industries must comply with specific regulations and rules. To the extent that AI predictive systems operate within a more highly regulated industry, variations in those laws may expand or otherwise impact the scope of liability for those products. The relevant concerns and risks present in each industry have been incorporated directly into the laws governing each of these sectors. For the purposes of this contribution, it is not possible to enumerate the many, highly complicated legal rules that have evolved along industry lines that may alter the liability risks incurred through the use of AI predictive systems. Instead, the goal here is merely to flag the AI tool's nature and its industry application as potential variables to be considered when assessing its liability risks.

II. DOES AI MERIT A NEW APPROACH TO LIABILITY?

From a liability perspective, where these technologies communicate and engage with one another, the potential culpability for anything that could go wrong shifts from one component to another as easily and quickly as the data transferred between them. Deciding who is responsible for what when something has gone wrong in a given case can be layered when many parties come into play. There are AI developers; algorithm trainers; data collectors, controllers, and processors; manufacturers of the devices incorporating the AI software; owners of the software (which are not necessarily the developers); and the final users of the devices (and perhaps many more related hands in the pot).

[25] *See, e.g.*, Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 902–13 (2018) (describing how transparency, complexity, and dimensionality issues combine to form an AI "black box"); Alex John London, *Artificial Intelligence and Black-Box Medical Decisions: Accuracy Versus Explainability*, HASTINGS CTR. REP., Jan.–Feb. 2019, at 15, 15–17 (2019).

[26] 1 STUART M. SPEISER ET AL., THE AMERICAN LAW OF TORTS § 1.3 (2013).

[27] *See id.* § 1.4.

[28] *See* 2A *id.* § 9:1; 5 *id.* § 18:1.

[29] *See, e.g.*, RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2(b) (AM. LAW INST. 1997).

Does this mean that AI merits a new approach to liability? A review of the literature suggests that there are roughly four possible answers to this question. A word of caution: it is not suggested that this is an exhaustive list of easily identifiable views but rather these possible answers operate on a continuum. Most approaches tend to edge closer to one category rather than the next depending on the context. The contribution that this Article offers is a snapshot of how the question above could be answered, free from normative implications.

A. Give AI Legal Personhood

In the complex AI Ecosystem, a plaintiff would need to establish whom to sue. If liability is hard to pin onto a particular tortfeasor (or group of tortfeasors), why could the AI itself not be held liable[30] if we were to treat AI systems as we do corporations?[31] In tort law, the concept of joint and several liability or the ability to sue anyone in the commercial chain for products liability claims suggests the likelihood of multiple possible defendants, especially as plaintiffs take into account the traditional interest in reaching a "deep pocket" capable of paying large damages. Absent use of comprehensive waivers, the AI Ecosystem will likely produce harms that will be litigated under traditional tort law. Whether traditional tort law and liability are the best ways to deal with modern technology is a more difficult question.

The idea of granting AI systems personhood is not science fiction but rather a legal fiction.[32] This approach rests on premises that would have to be explored further in a different venue. Suffice it to say that for it to work, it is necessary, at the very least, for the AI system to be capable of holding assets either directly (as a corporation would) or indirectly (assuming that either the licensor or the licensee of the AI system is to act on the AI system's behalf, as Jessica S. Allain suggests).[33] Unless the plaintiff can enforce a successful judgment to obtain redress for the loss suffered and that redress is most commonly financial, the purpose of granting legal personhood would be defeated in the first instance. For the purpose of this Article, the interesting aspect of the "AI legal personhood" line of reasoning is that, although it suggests a major change, it still aims at using existing legal rules to both advocate for and implement that change.

The liability risk associated with AI systems may vary depending on the nature of the AI—for example, whether the AI system is sold or licensed as

[30] *See generally* Gerhard Wagner, *Robot, Inc.: Personhood for Autonomous Systems?*, 88 FORDHAM L. REV. 591 (2019); Horst Eidenmüller, *The Rise of Robots and the Law of Humans* (Mar. 26, 2017) (unpublished manuscript), https:// papers.ssrn.com/sol3/ papers.cfm?abstract_id=2941001 [https://perma.cc/S7TT-GZBG].

[31] *Cf.* Jessica S. Allain, Comment, *From* Jeopardy! *to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems*, 73 LA. L. REV. 1049, 1078–79 (2013).

[32] *See, e.g.*, *id.* at 1078.

[33] *See id.*

a software or service or whether it is embedded in a tangible device.[34] The law often treats tangible items differently than intangible ones. For example, this is true for contractual liability, whereby the governing law may entirely change depending upon whether the AI is a "good" for sale.[35] The liability risk will also depend on the function of the AI output: is the system a predictive one on whose predictions humans base their decisions or is it fully autonomous such that the humans are "out of the loop,"[36] borrowing from the military terminology?[37]

AI systems are usually envisioned as making and executing decisions without human involvement. That is not necessarily so. Indeed, under European Union law, automated decisions that have legal or similar effects on individuals, as AI decisions may, are required to be subject to some type of human review.[38] That complicates the liability question by at least potentially putting another human being (or perhaps his or her legal person principal) at risk of liability.

Harm caused by an AI system could be the direct consequence of the AI's programming, perhaps giving rise to an intentional tort; negligent design, training, or operation (e.g., lack of adequate cybersecurity protections); or an arguably unforeseeable harm caused by an interaction with unforeseeable real-world data.

[34] AI systems are composed of algorithms and enabled by sophisticated software. Software is not necessarily considered a product but rather a service traditionally falling outside the reach of products liability. No case has yet applied strict products liability to software, though the Ninth Circuit has suggested, in dicta, that it may be possible to do so. *See* Winter v. G.P. Putnam's Sons, 938 F.2d 1033, 1035 (9th Cir. 1991). *See generally* RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 19 (AM. LAW INST. 1997).

[35] *See Mitigating Product Liability for Artificial Intelligence*, JONES DAY: INSIGHTS (Mar. 2018), https://www.jonesday.com/mitigating-product-liability-for-artificial-intelligence-03-21-2018 [https://perma.cc/27F4-RCNY].

[36] *See generally* PETER SINGER, WIRED FOR WAR: THE ROBOTICS REVOLUTION AND CONFLICT IN THE 21ST CENTURY (2009).

[37] There have been recent examples of highly sophisticated autonomous systems causing serious losses when they shut humans out of the loop. In the case of the Boeing 737 Max planes, it is alleged that the new Maneuvering Characteristics Augmentation System automatically pushed the plane's nose downward when the sensors detected that a stall might be imminent, causing the fatal crashes of Lion Air Flight 610 on October 29, 2018, and Ethiopian Airlines Flight 302 on March 10, 2019. *See, e.g.*, Anurag Kotoky, *Boeing's Grounded 737 Max—the Story so Far*, WASH. POST (July 5, 2019), https://www.washingtonpost.com/business/boeings-grounded-737-max-the-story-so-far/2019/07/04/bc7ed860-9e52-11e9-83e3-45fded8e8d2e_story.html [https://perma.cc/2X8Y-NQPH]. In March 2019, the engines of the Viking Sky cruise ship shut down in the middle of the tempestuous North Sea when, it is said, the sensors detected low oil level and forced the shutdown, leaving over one thousand passengers to be rescued by helicopter. *See, e.g.*, Lance Eliot, *Human In-the-Loop Vs. Out-of-the-Loop in AI Systems: The Case of AI Self-Driving Cars*, AI TRENDS (Apr. 9, 2019), https://www.aitrends.com/ai-insider/human-in-the-loop-vs-out-of-the-loop-in-ai-systems-the-case-of-ai-self-driving-cars/ [https://perma.cc/2DSR-R2CQ].

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

An individual's or organization's liability for tort stems directly from the potential tortfeasor's duty to avoid the harm that would result from the tortfeasor's conduct[39] or, as it has more traditionally been phrased, to avoid foreseeable harm.[40] To be held liable for tortious harm, that harm must be both the actual and legal result of the allegedly tortious act.[41] Given cases, such as *Palsgraf v. Long Island R. Co.*,[42] which show the doctrine of proximate or legal cause and manifest a limit to the duty of care, pure logical or actual causation is insufficient to establish tort liability.[43] Modern technology can raise substantial proximate cause issues.

Let us consider a simple hypothetical. Homeowner buys a "smart" smoke detector for her house. She buys the specific product because it has Wi-Fi capability which permits the vendor to send her a fire alarm notice to her smartphone if any of the sensors detect smoke or fire. She is particularly taken by the low cost of the internet-of-things product. A year later, she discovers that her home computer has been "hacked" and critical private banking data has been extracted and used in a way that greatly harmed her. The "hack" was made possible by the fact that the smoke detector was connected to her home Wi-Fi network, it had substandard cybersecurity protection (which is why it was so cheap), and the hacker penetrated her computer via the smoke detector.[44] During depositions, counsel learned that the defendant smoke alarm company had taken reasonable measures to protect its computer system but had not considered the detector itself to be a risk.

Rapid technological developments make it difficult to determine "risk," particularly in the price-sensitive commercial sector. The smoke detector example is a relatively simple one. Consider the more elaborate hypothetical often used by my William & Mary colleague, Professor Fred Lederer. Company is responsible for operating a dam and generating hydroelectric power. Company decides to modernize in order to be more efficient. It replaces its human-operated control system with a fully autonomous AI system. To enable the AI to function, Company installs a large number of sensors throughout the dam and the area in which the dam is. They collect temperature, moisture, stress, and other readings and send them via the internet to the AI. The "AI" actually consists of a number of components. The primary component is located in Company's primary corporate office some five hundred miles away. It constantly monitors the sensor data and

[39] RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL AND EMOTIONAL HARM § 29 (AM. LAW INST. 2005).

[40] *Id.* § 29 cmt. j.

[41] SPEISER ET AL., *supra* note 26, § 1:10.

[42] 162 N.E. 99 (N.Y. 1928) (involving a plaintiff who was injured at a busy train station by a heavy metal scale after it fell on her as an apparent result of train employees helping a passenger board a train, causing the passenger to drop a box of fireworks that exploded).

[43] 3 SPEISER ET AL., *supra* note 26, § 11:1.

[44] For a similar real case involving the penetration of a casino network via an aquarium sensor, see Alex Schiffer, *How a Fish Tank Helped Hack a Casino*, WASH. POST (July 21, 2017), https://www.washingtonpost.com/news/innovations/wp/2017/07/21/how-a-fish-tank-helped-hack-a-casino [https://perma.cc/2X5N-YUNW].

varies water flow on a continuous basis. It implements its decisions via instructions to its implementation module in the dam control room on site. Meanwhile the AI modifies its programming based upon its ongoing experience of the interaction of all the monitored sensor factors in order to produce the most electricity at the cheapest operating cost while maintaining community safety. The AI is also connected, via the internet, to other dam systems so that it can learn from how those systems are operating.

One night, the AI fully opens the emergency floodgates and floods one thousand homes downstream. Company investigates and cannot determine causation. Possibilities include: defective AI design; defective AI training; defective sensor design and/or manufacture; unforeseen consequences from multiple data inputs in real world circumstances; erroneous AI operation based upon sensor or remote data; and external interference, that could have been accidental or intentional, by either one or more private actors or on behalf of a foreign organization or nation. Notably, the sensors are from multiple companies and may have never been used together prior, certainly not in the instant configuration.

Causation and responsibility may be impossible to determine in the dam case. Every tort student's joy, res ipsa loquitur (the thing speaks for itself), which sometimes allows a court to conclude that a defendant must have been negligent because that is the only reasonable explanation for the harmful act,[45] cannot apply here. If nothing else, the possibility of external causation provides an alternative to system defects.[46]

Given multiple tortfeasors, American courts can usually apportion damages in a reasonable fashion.[47] That assumes, however, that the tortfeasors can be identified. Although traditional tort law provides a vehicle by which harm caused by the AI Ecosystem could potentially be compensated for, difficulties in proving causation may make that impossible. Critically, the difficulty in proving causation in these cases is "a feature and not a bug" and likely unavoidable in many cases.[48]

B. Leave AI Alone

The proponents of "leave AI alone" take a systemic view of the legal system, which is dynamic and capable of coping with the challenges posed by the so-called new technologies. Judge Frank Easterbrook famously said:

> [T]he best way to learn the law applicable to specialized endeavors is to study general rules. Lots of cases deal with sales of horses; others deal with people kicked by horses; still more deal with the licensing and racing of horses, or with the care veterinarians give to horses, or with prizes at horse shows. Any effort to collect these strands into a course on "The Law of the

[45] 2A SPEISER ET AL., *supra* note 26, § 9:1.

[46] Of course, in such a case, the lack of adequate cybersecurity may be the defect.

[47] *See, e.g.*, WARD FARNSWORTH & MARK F. GRADY, TORTS: CASES AND QUESTIONS 345–50 (2d ed. 2009).

[48] Edwina Rissland, *Artificial Intelligence and Law: Stepping Stones to a Model of Legal Reasoning*, 99 YALE L.J. 1957, 1962 (1990).

> Horse" is doomed to be shallow and to miss unifying principles. Teaching 100 percent of the cases on people kicked by horses will not convey the law of torts very well. Far better for most students—better, even, for those who plan to go into the horse trade—to take courses in property, torts, commercial transactions, and the like, adding to the diet of horse cases a smattering of transactions in cucumbers, cats, coal, and cribs. Only by putting the law of the horse in the context of broader rules about commercial endeavors could one really understand the *law* about horses.[49]

If an AI system makes a prediction or a decision, especially without substantial human involvement or oversight, the issue will be by what standard we should determine liability when unacceptable harm occurs but its causation cannot be determined.

The "leave AI alone" approach would remind us that the evolution of products liability law in the United States demonstrates that the legal system is capable of coping with difficulties in proving causation. Tort liability for harm caused by products applies even when the manufacturer proceeded reasonably and without negligence.[50] The only questions to be asked usually are: did the product cause harm, and, if so, was there a reasonable way to avoid that harm, even if we make that determination only through perfect hindsight?

"Self-driving" cars have raised the issue of AI liability most clearly. We want to encourage technological innovation while assisting those who may be hurt by the technology during and after the development period. Tort lawsuits could effectively stifle development. There are instances where traditional tort liability has been replaced either because it is substantially inefficient, such as in the case of no-fault auto insurance,[51] or outweighed by its potential negative consequences, as with the original September 11th Victim Compensation Fund—an administrative claims process to protect the airlines from potential massive liability.[52] In the case of self-driving cars, various jurisdictions have legislated or considered different approaches.[53] One arguably attractive solution is to create a compulsory no-fault quasi-insurance system in which the victim of a self-driving car receives a payment

[49] Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207, 207–08.

[50] *See, e.g.*, 5 SPEISER ET AL., *supra* note 26, § 18:2 n.16.

[51] *Background on: No-Fault Auto Insurance*, INS. INFO. INST. (Nov. 6, 2018), https://www.iii.org/article/background-on-no-fault-auto-insurance [https://perma.cc/Y4K2-65XP].

[52] Air Transportation Safety and System Stabilization Act, Pub. L. No. 107-42, 115 Stat. 230 (2001) (codified as amended in scattered sections of 49 U.S.C.). *See generally* KENNETH R. FEINBERG, WHAT IS LIFE WORTH?: THE UNPRECEDENTED EFFORT TO COMPENSATE THE VICTIMS OF 9/11 (2005).

[53] For an overview of federal and state legislation, see *Autonomous Vehicles: Self-Driving Vehicles Enacted Legislation*, NCSL (Sept. 4, 2019), http://www.ncsl.org/research/transportation/autonomous-vehicles-self-driving-vehicles-enacted-legislation.aspx [https://perma.cc/F4MB-L3B9] and *Autonomous Vehicles State Bill Tracking Database*, NCSL (Aug. 30, 2019), http://www.ncsl.org/research/transportation/autonomous-vehicles-legislative-database.aspx [https://perma.cc/HH33-QQFZ].

without the need to show why the automobile caused the harm.[54] The funds for the payment system would come from a surtax on each self-driving car.[55]

C. Robot Common Sense

Notably, a claims process assumes compensation and, in the case of unclear causation, compensation could be unfair. Or, compensation might slow or stop technological development. Accordingly, the most basic way to cope with AI Ecosystem harm may be to simply live with the harm as a necessary societal cost. There is precedent for such a result. In 2016, the Wisconsin Supreme Court decided *State v. Loomis*.[56] In *Loomis*, the criminal defendant was sentenced by a judge who took into account during sentencing the results of an AI predictive product, COMPAS.[57] The defense challenged the use of the AI tool and demanded access to the coding of its algorithm.[58] The trial judge denied the request on the basis of the trade secret evidentiary privilege.[59] The defendant complained inter alia of a due process violation because the sentencing court did not have accurate information, in part because the proprietary nature of COMPAS prevented him (and any other third party, including the court itself) from assessing its accuracy.[60] And in fact, ProPublica published a scathing criticism of COMPAS's accuracy a few months before the Wisconsin Supreme Court's decision was rendered.[61]

The Wisconsin Supreme Court upheld the use of the AI.[62] In large part, the opinion is based on the fact that it was the judge who sentenced the defendant—not the AI. The COMPAS prediction was only one factor,[63] and the defendant had access to the questions used to produce the data given to the AI.[64] But, there is an undertone in the opinion which suggests that another reason was important: human sentencing is filled with flaws.[65]

[54] *Cf.* Giuffrida, Lederer & Vermeys, *supra* note 21, at 765.

[55] *See id.* at 764 n.58.

[56] 881 N.W.2d 749 (Wis. 2016).

[57] COMPAS stands for Correctional Offender Management Profiling for Alternative Sanctions, and it is a relatively common criminal risk assessment tool. Equivant, the entity developing "software for justice," states that COMPAS is a management support tool that "helps inform your critical decisions and mitigate . . . risk." *Classification Module*, EQUIVANT, https://www.equivant.com/compas-classification [https://perma.cc/BJ24-PU8P] (last visited Oct. 6, 2019).

[58] Brief of Defendant-Appellant at 22–25, *Loomis*, 881 N.W.2d 749 (No. 2015AP000157-CR).

[59] *Id.* at 3. The complexity of AI coding, training, and operation gives rise to the movement for "Explainable AI." In *Loomis*, the evidentiary privilege mooted even the possibility of examining the code to better understand its results. *See Loomis*, 881 N.W.2d at 753.

[60] *Id.* at 757.

[61] Julia Angwin et al., *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/4RVR-KD5A].

[62] *Loomis*, 881 N.W.2d at 772.

[63] *Id.* at 753, 767–68, 771.

[64] *Id.* at 761–62.

[65] *See id.* at 765.

Technology-augmented sentencing, however erroneous, has at least the possibility of improving over time and curing the current—and defective—human system.[66] In short, we should live with arguably inadequate AI because in the long term its descendants may be better than what we otherwise would use.

The *Loomis* decision is an illustration of how "robot common sense" is making its way into the legal system. With "robot common sense," we refer not only to the legal system adapting concepts such as reasonableness and foreseeability to capture the nuances brought in by AI and new technologies but also to our expectation that an AI system will update its common sense by learning from external input.[67] Perhaps machine learning is more robust and less fallible than human learning, and "robot common sense" is preferable to the reasonable person standard. Or perhaps not—at least not in the criminal justice context.[68]

D. New (Non-Legal) Rules for AI: Is Code Law?

If we are to use a compensation-oriented system for AI Ecosystem harms, a more general approach may be useful. In their 2019 article "Remedies for Robots," Mark Lemley and Bryan Casey suggest a harms-based approach.[69] Potential harms include unavoidable or inherent harms; deliberate or least-cost harms; defect-driven harms; misuse harms; unforeseen harms; systemic harms; and collateral harms.[70]

In this line of reasoning, the compensation-deterrence methodology could be harm-specific, rather than focusing on the tortfeasor(s). If self-driving cars contain unavoidable risks, no-fault compensation may be appropriate. Ascertained intentional least-cost harms may merit a lawsuit via some form of products liability approach.

Starting with specific risks of harm has been a favorite strategy within more developed areas of technology regulation in the United States. For example, within the fields of data privacy and cybersecurity, federal standards have migrated towards a risk-based approach, allowing flexibility for data controllers and processors to determine their individual risk and risk

[66] This, however, presupposes that the AI is learning from either fresh data or supervised learning rather than, as some worry, crystallizing the bias and repeating it ad infinitum. *See, e.g.*, Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677–93 (2016); Kate Crawford, *The Hidden Biases in Big Data*, HARV. BUS. REV. (Apr. 1, 2013), https://hbr.org/2013/04/the-hidden-biases-in-big-data [https://perma.cc/EQQ8-ZQU6].

[67] Curtis E. A. Karnow, *The Application of Traditional Tort Theory to Embodied Machine Intelligence*, *in* ROBOT LAW 51, 53–61 (Ryan Calo et al. eds., 2016); Harry Surden & Mary-Anne Williams, *Technological Opacity, Predictability, and Self-Driving Cars*, 38 CARDOZO L. REV. 121, 147–48 (2016).

[68] *See generally* Anne Washington, *How to Argue with an Algorithm: Lessons from the COMPAS-ProPublica Debate*, 17 COLO. TECH. L.J. 131 (2018).

[69] *See generally* Mark A. Lemley & Bryan Casey, *Remedies for Robots*, 86 U. CHI. L. REV. 1311 (2019).

[70] *Id.* at 1327–42.

tolerance on a case-by-case basis.[71] Similar approaches have been suggested for AI,[72] as the nature and degree of the risks associated with the technology will require different governance and will potentially lead to divergent liability outcomes. It is therefore important to decide, when AI intersects with these areas of law, which harms we are looking to avoid and who is best able to prevent the harms (and thus who should be held accountable for failing to do so).[73]

Theories of liability, along with legal and equitable remedies, are often founded on a certain belief about human motivation—the innate desire to avoid punishment. These remedies are arguably ill-suited for AI systems, and thus do not carry the same weight in shaping AI's behavior (assuming behavior is even the right term). To apply the same approach to AI, Lemley and Casey suggest that the incentives would need to be coded, according to relative weights, directly into the algorithms, which is a difficult feat.[74] There are significant ethical and philosophical considerations behind assigning numeric significance to competing priorities, not to mention the practical problem of effectively coding this incentive system into more sophisticated machine learning apparatuses. Not all remedies are designed solely to affect behavior, and, often, the ability to shape and control harm is limited by which theory of liability is applied. Harm-based remedies,[75] as opposed to those based on the idea of human desire to avoid punishment,[76] may more appropriately apply in the AI context. In light of the above practical challenges, companies that are able to internalize the costs of potential liability may elect to do so, rather than avoid the behavior. The likelihood of successful claims may well be low, making the likelihood of

[71] *See, e.g.*, Federal Information Security Modernization Act of 2014, 44 U.S.C. § 3553(a)(4) (Supp. 2014); NAT'L INST. OF STANDARDS & TECH., FRAMEWORK FOR IMPROVING CRITICAL INFRASTRUCTURE CYBERSECURITY 1 (2018), https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf [https://perma.cc/RG2P-PZ2H]. However, some states have taken a different approach. For example, under Illinois's Biometric Information Privacy Act, 740 ILL. COMP. STAT. 14 (2019), the Illinois Supreme Court held that plaintiffs are not required to show that they suffered harm other than a violation of the law in order to launch proceedings. Rosenbach v. Six Flags Entm't Corp., 129 N.E.3d 1197, 1207 (Ill. 2019).

[72] *See generally* Antje von Ungern-Sternberg, *Artificial Agents and General Principles of Law*, 60 GERMAN Y.B. INT'L L. 239 (2017); Wagner, *supra* note 30.

[73] *See* sources cited *supra* note 72.

[74] *See generally* Lemley & Casey, *supra* note 69. For the idea of "code is law," see generally LAWRENCE LESSIG, CODE: VERSION 2.0 (2006) and Lawrence Lessig, *The Law of the Horse: What Cyberspace Might Teach*, 113 HARV. L. REV. 501 (1999).

[75] *See, e.g.*, *Consequential Damages*, BLACK'S LAW DICTIONARY (11th ed. 2019) ("Losses that do not flow directly and immediately from an injurious act but that result indirectly from the act."); *Expectation Damages*, BLACK'S LAW DICTIONARY (11th ed. 2019) ("Compensation awarded for the loss of what a person reasonably anticipated from a transaction that was not completed.").

[76] *See Punitive Damages*, BLACK'S LAW DICTIONARY (11th ed. 2019) ("Damages awarded in addition to actual damages when the defendant acted with recklessness, malice, or deceit . . . .").

less-costly settlements far greater.[77] If the popularity of and the high demand for AI products continue, it may simply become worth the risk.

Risk-based, least-cost avoidance liability approaches have also attracted a fair number of critics.[78] They argue that the liability regime predominant in law creates, or at least preserves, the status quo for a system of profit maximization.[79] In other words, it creates a floor for avoiding the worst behaviors but not a ceiling of best uses to aspire towards. There is incentive to avoid the worst errors but perhaps not enough to entice companies to innovate and create the best possible results. What is more, where injury seems likely to occur, companies may choose to simply accept the risk and externalize the predicted costs of liability by increasing consumer prices. In this case, whatever incentive structure may have existed to avoid harm through financial motivations is eliminated.[80]

If we use regulation rather than tort suits, regulators will have to decide optimal risk tolerances: how much harm are we willing to allow to obtain the social benefits of AI? More lenient liability rules will allow more harms to go uncompensated, but this in turn incentivizes more—and especially smaller—developers to innovate and potentially create better AI systems without the fear of large lawsuits. Some scholars are, however, quick to point out that liability rules do not need to be one-size-fits-all, and the severity of the harms caused can be a major determining factor for when harsher or more lenient liability rules apply.[81]

Whether we are to use current tort law, claims processes, or regulatory devices, we must unavoidably deal with a cost-benefit analysis. If we adapt existing legal frameworks, our benefits are: (1) that we need only adapt current law, which we are used to doing; and (2) that common law allows case-by-case resolution—even when dealing with unexpected matters. The arguable drawbacks are: (1) the likely unintended legal consequences which may impede technology development or business; (2) the gaps until new harms or new cases arise; (3) the rules that are likely to vary greatly in their speed of development and sufficiency, depending upon the legal and economic sector involved; and (4) the legal uncertainty that discourages capital investment and use of novel science and technology.

However, adopting new legal frameworks inevitably involves substantial error while we cope with the pragmatic effects of untried rules or procedures—especially as technology changes constantly.

[77] Bryan Casey, *Amoral Machines, or: How Roboticists Can Learn to Stop Worrying and Love the Law*, 111 NW. U. L. REV. ONLINE 231, 242–46 (2017).

[78] For a critical analysis of the literature, see generally Stephen G. Gilles, *Negligence, Strict Liability, and the Cheapest Cost-Avoider*, 78 VA. L. REV. 1291 (1992).

[79] *See id.* at 1350.

[80] Casey, *supra* note 77, at 242–44.

[81] *See generally* Lemley & Casey, *supra* note 69.

III. AI ADVANCES: AN "ESCALATOR FROM HELL"?[82]

Determining how legal rules should apply to liability for losses caused by AI systems will take time, but in the interim AI is making strides in everyday life. The medical profession has enshrined the duty to "do no harm."[83] Perhaps that should be the critical organizing principle in these early days of the AI Ecosystem. Rather than attempting to adapt or create a full compensatory system for AI harms, we should instead focus on identifying and dissuading (and perhaps compensating) the major predictive harms, with the understanding that constant reevaluation will be necessary. In doing so, we would therefore be wise to contemplate the emerging ethical codes for AI systems.

While considering the morality of machines may be unnecessary, if not futile, at this stage,[84] those who use AI systems are fallible, corruptible, imperfect human beings. From intentional misuses of AI systems[85] to unintentional consequences,[86] developing AI cannot be done in the abstract. As Sir Robin Knowles, an English High Court judge, said recently, an ethical and legal framework is "imperative" for AI.[87] He is not alone in calling for ethics to be embedded in the regulation of AI.[88]

Efforts have already been made. The "trolley problem" has been adapted to an AI-augmented reality because AI systems have life-or-death

[82] Alex Hern, *New AI Fake Text Generator May Be Too Dangerous to Release, Say Creators*, GUARDIAN (Feb. 14, 2019), https://www.theguardian.com/technology/2019/feb/14/elon-musk-backed-ai-writes-convincing-news-fiction [https://perma.cc/AUJ2-8W33] (quoting Jack Clark, the policy director of OpenAI).

[83] 1 STEVEN E. PEGALIS, AMERICAN LAW OF MEDICAL MALPRACTICE § 3:8 (3d ed. 2019).

[84] *See generally* Casey, *supra* note 77.

[85] Examples abound, but see, for instance, "deep fakes," which are "hyper-realistic digital falsification of images, video, and audio." *See generally* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. (forthcoming 2019), https://ssrn.com/abstract=3213954 [https://perma.cc/9HKQ-ABZX]. In February 2019, OpenAI refused to make available its newest AI model, GPT2, because OpenAI considered GPT2's capabilities so strong that the risk of malicious and nefarious uses outweighed the benefit of sharing the results of its research for an open and informed discussion about practical as well as ethical implications. *See* Hern, *supra* note 82.

[86] Think, for instance, of unconscious bias in AI recruitment systems. *See, e.g.*, Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, REUTERS (Oct. 9, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [https://perma.cc/44EK-3ABF].

[87] Nick Hilborne, *High Court Judge: Ethical and Legal Framework for AI "Imperative*,*"* LEGAL FUTURES (Nov. 13, 2018), https://www.legalfutures.co.uk/latest-news/high-court-judge-ethical-and-legal-framework-for-ai-imperative [https://perma.cc/27FQ-P3YU].

[88] *See generally* Luciano Floridi et al., *AI4People—an Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*, 28 MINDS & MACHINES 689 (2018).

implications[89]: self-driving cars[90] are on our roads and accidents are already occurring.[91] How will a self-driving car act when harm is unavoidable? The importance of ethics in the discourse surrounding the growth of AI systems both in terms of uses and capabilities cannot be underestimated. However, space constraints dictate that only few considerations can be put forward in this Article.

Ethical codes are starting to emerge, and not only from academic circles. In December 2018, the European Commission for the Efficiency of Justice of the Council of Europe[92] adopted the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems[93] (the "Charter"). A few days later, the High-Level Expert Group on AI appointed by the European Commission published the first draft of its Ethics Guidelines for Trustworthy AI, the final version of which was released in mid-2019.[94] These guidelines make it clear that developments in AI do not take place in a lawless vacuum but under the rule of law of the European Union (e.g., the Charter and the General Data Protection Regulation) and international law (e.g., U.N. human rights treaties).[95] The report goes so far as to propose "[a] future where democracy, the rule of law and fundamental rights underpin AI systems."[96] The report also offers four key principles underpinning the development of AI systems: respect for human autonomy, prevention of harm, fairness, and

[89] The "trolley problem" is a thought experiment developed by ethicists and moral philosophers to highlight the moral and ethical challenges of binary choices involving life-and-death decisions. Kyle Wiggers, *MIT Study Explores the "Trolley Problem" and Self-Driving Cars*, VENTUREBEAT (Oct. 24, 2018, 4:40 PM), https://venturebeat.com/2018/10/24/mit-study-explores-the-trolley-problem-and-self-driving-cars [https://perma.cc/J55G-KERC].

[90] Although not yet fully autonomous, self-driving cars have many autonomous features. Doug DeMuro, *7 Best Semi-Autonomous Systems Available Right Now*, AUTOTRADER (Jan. 2018), https://www.autotrader.com/best-cars/7-best-semi-autonomous-systems-available-right-now-271865 [https://perma.cc/U8MM-DDWK].

[91] Daisuke Wakabayashi, *Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam*, N.Y. TIMES (Mar. 19, 2018), https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html [https://perma.cc/3NUV-D6PM].

[92] The Council of Europe is an international organization not to be confused with the European Union.

[93] European Comm'n for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment*, COUNCIL EUR., https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c [https://perma.cc/U8F7-N6VA] (last visited Oct. 6, 2019). Key principles include respect for fundamental rights by compatible design and implementation of AI tools and services; nondiscrimination, "specifically prevent[ing] the development or intensification of any discrimination between individuals or groups of individuals"; quality and security for "the processing of judicial decisions and data"; "transparency, impartiality and fairness" through "data processing methods accessible and understandable," and external audits; and "preclud[ing] a prescriptive approach and ensur[ing] that users are informed actors and in control of the choices made." *Id.* at 7.

[94] *See* High-Level Expert Grp. on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* (Apr. 8, 2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419 [https://perma.cc/L3JA-TR7B].

[95] *Id.* at 6.

[96] *Id.* at 9.

explicability.[97] Notably, it does not attempt to limit what developers can do with AI. Rather, it flags areas of "critical AI concern," the legal ramifications of which are not yet fully understood.[98]

More recently, the People's Republic of China (PRC) has put forward the foundation of its own vision for ethics in AI which, unsurprisingly, has a different tenor from the European approach. In May 2019, the Beijing Academy of Artificial Intelligence[99] (BAAI) released the Beijing AI Principles,[100] which were followed in July 2019 by the Governance Principles for the New Generation Artificial Intelligence, published by the National Governance Committee for the New Generation Artificial Intelligence (the "Governance Principles").[101] Interestingly, the Beijing AI Principles appear to suggest that informed consent must be sought by AI developers before individuals use their products,[102] opening the door for potential data protection measures that echo the European ones. There is even a statement, if the translation is to be trusted, that "[h]uman privacy, dignity, freedom, autonomy, and rights should be sufficiently respected. AI should not be used to against, utilize or [sic] harm human beings."[103] The Governance Principles make a timid reference to respecting human rights, but subject to "social security" (a peaceful, law-abiding society), and they include a reference for the need for AI-users to be informed of potential risks.[104]

However, the overarching objectives that the PRC's ethical vision pushes forward are aligned with its goal of becoming the leading AI power by 2030. As James Lomonosoff puts it, this reveals the Chinese government's security-first approach, which goes hand-in-hand with the export of the surveillance state;[105] this is in contrast with the humanity-first approach of the European Union, which, although linked to notions of rule of law and democracy, has not yet put forward any legal frameworks.

[97] *Id.* at 12.

[98] *Id.* at 33; *see* James Lomonosoff, *The European Union and China: Two Differing Approaches to Ethics in Artificial Intelligence*, CTR. FOR LEGAL & CT. TECH. (2019), https://legaltechcenter.openum.ca/files/sites/159/2019/08/The-EU-and-China_Two-Differing-Approaches-to-Ethics-in-AI.pdf [https://perma.cc/9MVQ-VEQ8].

[99] The BAAI is one of the many initiatives funded by the Chinese government to meet the ambitious goal of becoming the world's foremost AI innovation center by 2030. *See generally* JEFFREY DING, DECIPHERING CHINA'S AI DREAM: THE CONTEXT, COMPONENTS, CAPABILITIES, AND CONSEQUENCES OF CHINA'S STRATEGY TO LEAD THE WORLD IN AI 10 (2018), https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf [https://perma.cc/DG26-YTZJ].

[100] *Beijing AI Principles*, BEIJING ACAD. ARTIFICIAL INTELLIGENCE (May 28, 2019), http://www.baai.ac.cn/blog/beijing-ai-principles [https://perma.cc/8VLR-WABM].

[101] *Governance Principles for the New Generation Artificial Intelligence—Developing Responsible Artificial Intelligence*, CHINA DAILY (June 17, 2019, 3:59 PM), www.chinadaily.com.cn/a/201905/17/WS5d07486ba3103dbf14328ab7.html [https://perma.cc/LF9G-ZMPJ].

[102] *Beijing AI Principles*, *supra* note 100.

[103] *Id.*

[104] *Governance Principles for the New Generation Artificial Intelligence*, *supra* note 101.

[105] Lomonosoff, *supra* note 98.

The United States has taken a different approach. President Trump's Executive Order 13,859 clearly encourages the development of American AI but does not make any reference to ethics.[106] With the introduction of the Algorithmic Accountability Act of 2019,[107] which focuses on bias in AI decision-making, Congress has shown an interest in regulating aspects of AI, but without a commitment to particular ethical values. Thus, it is outside the formal rulemaking power that ethical values are being discussed. Private entities, especially Big Tech, are putting forward their visions of how their normative AI systems should be used.[108] Last summer, Google did not seek an extension of a high-value contract with the U.S. Department of Defense as a consequence of pressure from its employees concerned that their AI advances would be deployed for military uses.[109] The American approach is, therefore, a privately driven one. Whether this is preferable and more inclusive than state-centric approaches remains to be seen.

CONCLUSION

Liability for losses linked to AI systems is a difficult emerging area that requires further analysis. Some form of regulation is inevitable and certainly desirable, but there is a pressing need to address the core ethical questions that AI systems pose. While international powers position themselves for potential hegemonic control of, and through, AI, the scientific community, the private sector, and think tanks are pushing for more transparent and explainable AI systems. As shown by the breadth and depth of the presentations at the excellent *Rise of the Machines* Symposium, the best models that will help us frame the right questions and formulate the proper answers to AI challenges come from interdisciplinary efforts.

[106] Exec. Order No. 13,859, 84 Fed. Reg. 3967 (Feb. 11, 2019).

[107] H.R. 2231, 116th Cong. (2019).

[108] *See, e.g.*, *Frequently Asked Questions*, PARTNERSHIP ON AI, https://www.partnershiponai.org/faq [https://perma.cc/MV2M-JSWK] (last visited Oct. 6, 2019) ("The Partnership on AI (PAI) is a multistakeholder organization that brings together academics, researchers, civil society organizations, companies building and utilizing AI technology, and other groups working to better understand AI's impacts."). Partnership on AI has over ninety partners, ranging from Apple to Facebook to Google, across thirteen countries. *See also Artificial Intelligence at Google: Our Principles*, GOOGLE AI, https://ai.google/principles [https://perma.cc/32VC-49AC] (last visited Oct. 6, 2019); *Microsoft AI Principles*, MICROSOFT, https://www.microsoft.com/en-us/ai/our-approach-to-ai [https://perma.cc/3Z2R-UDT7] (last visited Oct. 6, 2019).

[109] Drew Harwell, *Google to Drop Pentagon AI Contract After Employee Objections to the "Business of War*,*"* WASH. POST (June 1, 2018), https://www.washingtonpost.com/news/the-switch/wp/2018/06/01/google-to-drop-pentagon-ai-contract-after-employees-called-it-the-business-of-war [https://perma.cc/X85C-VPUW].