Yale Journal of Law & Technology
Volume 26, Issue 2

Humans Outside the Loop

Charlotte A. Tschider[*]

Artificial Intelligence (AI) is not all artificial. Despite the need for high-powered machines that can create complex algorithms and routinely improve them, humans are instrumental in every step used to create AI. From data selection, decisional design, training, testing, and tuning to managing AI's development as it is used in the human world, humans exert agency and control over the choices and practices underlying AI products. AI is now ubiquitous: it is part of every sector of the economy and many people's everyday lives. When AI development companies create unsafe products, however, we might be surprised to discover that very few legal options exist to remedy any wrongs.

This Article introduces the myriad of choices humans make to create safe and effective AI products and explores key issues in existing liability models. Significant issues in negligence and products liability schemes, including contractual limitations on liability, distance the organizations creating AI products from the actual harm they cause, obscure the origin of issues relating to the harm, and reduce the likelihood of plaintiff recovery. Principally, AI offers a unique vantage point for analyzing the limits of tort law, challenging long-held divisions and theoretical constructs. From the perspectives of both businesses licensing AI and AI users, this Article identifies key impediments to realizing tort doctrine's goals and proposes an alternative regulatory

---

[*] Associate Professor, Loyola University Chicago School of Law. I would like to thank Matthew Sag, Michael Madison, Ryan Calo, Roger Ford, Julie Cohen, James Grimmelman, Tabrez Ebrahim, Salomé Viljoen, Sari Mazzurco, Peter Lee, Peter Salib, the participants of the Data in Business and Society Lewis & Clark Law Fall Forum, and attendees of the Future of Law in Technology and Governance (Future Law Project) at the University of Pittsburgh's Center for Governance and Markets for their excellent suggestions. I would also like to thank my wonderful research assistant, Natasha Ganesh, for her commitment to this and a number of other simultaneous projects, and to the *YJoLT* editors for their excellent work getting this piece to its final form.

scheme that shifts liability from humans in the loop to humans outside the loop.

Article Contents

Introduction

Artificial Intelligence (AI) is part of nearly every sphere of modern society.[1] AI is integrated into critical infrastructure, agriculture, transportation, manufacturing, healthcare, finance, and consumer goods.[2] While only 33% of consumers think they are using AI platforms, at least 77% are actually using them, and at least 77% of devices currently feature some form of AI.[3] To date, 35% of businesses have adopted a version of AI.[4] AI can tell you when you are going to have a migraine, when it is time to invest in a new stock, or when you are about to get into a car accident.[5] AI is designed to assess, diagnose, recommend, alert, and automate physical function—it is positioned to leverage complex decisional systems to overcome human problems and improve a human world.[6] AI's artificiality, its distinction from human decision-making, powers solutions to intractable human problems.[7]

---

[1] Rebecca Crootof, Margot E. Kaminski & W. Nicholson Price II, *Humans in the Loop*, 76 VAND. L. REV. 429, 432-33 (2023).
[2] *Id.*
[3] Mark Webster, *149 AI Statistics: The Present and Future of AI at Your Fingertips,* AUTHORITYHACKER (Aug. 7, 2023), https://www.authorityhacker.com/ai-statistics [https://perma.cc/N8PK-7C7Z].
[4] *Id.*
[5] Daniel Riskin et al., *Using Artificial Intelligence to Identify Patients with Migraine and Associated Symptoms and Conditions with Electronic Health Records*, 23 BMC MED. INFORMATICS & DECISION MAKING, July 14, 2023, at 1, 2; Andrew Wilson, *Can AI Predict the Stock Market? Yes, and Here's How*, MEDIUM (Oct. 5, 2023), https://medium.com/@TheAndrewLab/can-ai-predict-the-stock-market-yes-and-heres-how-d2486cffc6ff [https://perma.cc/MFM3-6FSL]; Rachel Gordon, *Deep Learning Helps Predict Traffic Crashes Before They Happen*, MIT NEWS (Oct. 12, 2021), https://news.mit.edu/2021/deep-learning-helps-predict-traffic-crashes-1012 [https://perma.cc/V98P-5FUH].
[6] Minkyu Shin, Jin Kim, Bas van Opheusden & Thomas L. Griffiths, *Superhuman Artificial Intelligence Can Improve Human Decision-Making by Increasing Novelty*, 120 PNAS art. no e221484020, at 1 (Mar. 13, 2023).
[7] *Id.*

AI is not *completely* artificial, though. Human data scientists, at least initially, design, train, and test AI.[8] Humans, through their interactions with technology, produce the data used to train AI.[9] Humans use AI, potentially revealing how and to what extent AI operates in socially undesirable ways.[10] AI will not function effectively without humans involved in its design and operation.[11] However, despite the effectiveness of human contributions in creating and using AI, society should be cautious when expecting humans to challenge, interrupt, or supervise AI.

Although AI has been long imagined, at least since Alan Turing's famous Turing Test,[12] humanity is in its infancy with respect to complex AI, such as deep learning applications designed to solve our most complex problems. Our lack of confidence in AI's safety, effectiveness, and fairness, coupled with fears related to its opacity, has prompted calls for AI oversight.[13] Absent concrete regulatory interventions, however, AI developers are largely left to their own devices to determine how to create products that are unlikely to result in harm to those who use them.[14]

---

[8] Joe McKendrick & Andy Thurai, *AI Isn't Ready to Make Unsupervised Decisions*, HARV. BUS. REV. (Sept. 15, 2022), https://hbr.org/2022/09/ai-isnt-ready-to-make-unsupervised-decisions [https://perma.cc/YW2R-XJYR].

[9] *Id.*

[10] *Id.*

[11] *Id.*

[12] The universal computer, or the Universal Turing Machine, anticipated continuous learning by permitting controls or rules to be updated like data, using a meta-logical interpreter. This continuous learning capacity is key for AI in differentiating between human-designed algorithms and self-executing AI. The Turing Test, therefore, is the inability to differentiate between a human and a computer, supported by the Universal Turing Machine. Alan M. Turing, *Computing Machinery and Intelligence*, 49 MIND 433, 439 (1950).

[13] Crootof et al., *supra* note 1, at 445.

[14] The Biden Administration's AI Executive Order evidences this approach as well, largely relying upon AI organizations' commitments to self-regulation. Exec. Order No. 14110 (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/presidential-

Modern AI is a combination of system design and architecture, data selection and organization, training practices and performance criteria.[15] But like any other type of technology designed by humans, AI can fail to perform as expected due to human mistakes and failure to anticipate potential risks.[16] AI may injure humans or cause property damage.[17] It could compromise individual privacy or perpetuate and entrench discrimination. AI decisions may influence human knowledge, attitudes, and behavior.[18] However, AI's black box nature often means that its decisions will not be readily intelligible, even when examined by the AI's designers.[19]

For these reasons, humans "in the loop" have been positioned as objective yet curative intervenors who can prevent AI harm from occurring and, if harm occurs, shield AI

---

actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence [https://perma.cc/K5S6-T79M]. The National Institute on Standards and Technology (NIST) issued a request for comment on the executive order, which closed on February 2, 2024. *NIST's Responsibilities Under the October 30, 2023 Executive Order*, NAT'L INST. STANDARDS & TECH. (2023), https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence [https://perma.cc/NZK6-6DHF].

[15] Sven Blumberg, Jorge Machado, Henning Soller & Asin Tavakoli, *Breaking Through Data-Architecture Gridlock to Scale AI*, MCKINSEY DIGITAL (Jan. 26, 2021), https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/breaking-through-data-architecture-gridlock-to-scale-ai [https://perma.cc/7SJM-NXQJ].

[16] Jorge Constantino, *Exploring Article 14 of the EU AI Proposal: Accountability Challenges of the Human in the Loop When Supervising High-Risk AI Systems in Public Administration*, VU UNIVERSITY AMSTERDAM LEGAL STUDIES PAPER SERIES 4, 23 (2022).

[17] *Id.*; Rowena Rodrigues, *Legal and Human Rights Issues of AI: Gaps, Challenges and Vulnerabilities*, 4 J. RESPONSIBLE TECH., Oct. 16, 2020, at 1, 5.

[18] Bobby Hoffman, *The Hidden Mental Manipulation of Generative AI*, PSYCH. TODAY (Aug. 3, 2023), https://www.psychologytoday.com/us/blog/motivate/202307/the-hidden-mental-manipulation-of-generative-ai?eml [https://perma.cc/AW5B-Z4RF].

[19] McKendrick & Thurai, *supra* note 8; Saurabh Bagchi, *What Is an AI Black Box? A Computer Scientist Explains*, HOWSTUFFWORKS (May 26, 2023), https://computer.howstuffworks.com/ai-black-box.htm [https://perma.cc/3C94-UBUB].

creators from liability.[20] First, a human in the loop could be essential to the function of some AI applications that are designed to work alongside a human operator or user.[21] Second, a human in the loop may be placed in the loop to mitigate risk, supervising AI's function and intervening when it malfunctions.[22] Third, a human in the loop may be used strategically as legal protection, absorbing or shifting liability from an AI creator or reducing risk and corresponding regulatory obligations.[23] In its most damaging form, a human in the loop shields the creator from liability when someone is harmed, functionally preventing the real examination of unsafe, ineffective, or unfair technology.[24]

AI creators benefit from liability models and contractual private ordering (including limitations on liability) that foreclose recovery by injured parties. While these issues are not new, AI makes them more complex. Plaintiffs may have difficulty meeting prima facie case requirements when AI is involved. Defendants to a variety of tort claims, such as products-liability actions, have long used human error as an affirmative defense through comparative or contributory negligence; humans in the loop broaden the circumstances where this strategy can be used.[25] Similarly, contractual limits

---

[20] Crootof et al., *supra* note 1, at 473-75. Crootof, Kaminski, and Price define a human in the loop as "an individual who is involved in a single, particular decision made in conjunction with an algorithm." *Id.* at 440. See *infra* Part I for an expanded discussion what the loop entails and which actors are in or outside of it.

[21] *See, e.g.*, Tamara Scott, *11 Examples of Humans and AI Working Together in Small Business*, SMALL BUS. COMPUTING (Oct. 7, 2021), https://www.smallbusinesscomputing.com/it-management/humans-ai-working-together-small-business [https://perma.cc/N62W-JUE5] (describing the variety of AI people work with today).

[22] *Id.* at 474-78.

[23] *Id.* at 482.

[24] *See infra* Parts II and III. *See* Pegah Moradi & Karen Levy, *The Future of Work in the Age of AI: Displacement or Risk-Shifting?*, *in* THE OXFORD HANDBOOK OF ETHICS OF AI 270, 278-79 (2020).

[25] States have adopted different liability models, for example contributory negligence is used in a minority of jurisdictions and relieves injury if the

on liability, liability-shifting provisions, indemnification, and disclaimed implied warranties have been used for a variety of products and services to change liability outcomes through private agreement.[26] These methods are now being used to require consumers to act as humans in the loop and supervise AI they do not and cannot understand. Cognitive challenges to understanding technology, such as automation bias and information asymmetry between AI creator and AI user, may paradoxically make consumers more trustful of AI decisions while limiting their understanding of AI.

These challenges stalemate the development of appropriate interventions for ensuring AI is safe, effective, and fair.[27] Although many commentators have noted that AI is not safe enough to function without supervision,[28] positioning humans as supervisors is not a solution to AI's safety concerns.[29] Rather than focus on improving safety, humans have *become* the safety in otherwise unsafe systems.[30]

---

plaintiff is even partially responsible for their injury. *See Contributory and Comparative Negligence by State*, BLOOMBERG L., (Jan. 2023), https://pro.bloomberglaw.com/brief/contributory-and-comparative-negligence-by-state [https://perma.cc/GJ5N-TMVK]. Comparative negligence apportions liability by percentage. Some states include modified rules—for example, if one party is over 50% at fault, they may inherit all of the liability. *Id.*

[26] Scott Walters, *Risk Shifting Clauses: Is a Trend Emerging?*, SMITH CURRIE, https://www.smithcurrie.com/publications/common-sense-contract-law/risk-shifting-clauses-is-a-trend-emerging/pdf [https://perma.cc/4JGU-UQG8]; Richard H. Casper & Kathleen E. Wegrzyn, *Commercial and Consumer Warranties: A Primer*, FOLEY & LARDNER LLP (Mar. 20, 2019), https://www.foley.com/en/insights/publications/2019/03/commercial-and-consumer-warranties-a-primer [https://perma.cc/7VC4-NMXR].

[27] *See infra* note 74.

[28] Joe McKendrick & Andy Thurai, *supra* note 8; John Davis Akkara & Anju Kuriakose, *Commentary: Is Human Supervision Needed for Artificial Intelligence?*, 70 INDIAN J. OPHTHALMOLOGY 1138, 1138 (2022).

[29] Bill Franks, *Human in the Loop Is a Popular Way to Mitigate the Risks of AI. That Approach Might Be Doomed*, MEDIUM (Sept. 7, 2021), https://medium.com/analytics-matters/human-in-the-loop-is-a-popular-way-to-mitigate-the-risks-of-ai-that-approach-might-be-doomed-f6631cd1e78d [https://perma.cc/BF8S-KETA].

[30] Part of the concern is that, in computerized systems, a human could make the situation worse or better based on the contextual scenario and what they

This Article proceeds in three parts. Part I introduces AI principles and examines AI's technical foundations, explaining where humans *outside* the loop are essential to the development of safe and accurate technologies. Part II explores the various liability models that likely apply to AI and the limitations of placing humans *inside* the loop, identifying key gaps and inconsistencies in tort law that ultimately interfere with its goals, leaving plaintiffs without compensation and rewarding behaviors that are inconsistent with public policy goals of safe, effective, and fair AI.[31] Part III offers a starting point for reconceptualizing a statutory and liability model that rewards certain behavior to create safer, more effective, and fairer AI without increasing risk for the people least able to mitigate or avoid it.

I. The "Loop"

Before we examine how humans in the loop interfere with AI safety, efficacy, and fairness goals, we must understand how AI is designed and developed. The "loop," so to speak, is the AI decisional process from start to end, a cyclic process where the interactions with humans in prior loops serve as inputs for loops that follow.[32] Unlike a human-designed algorithm that makes a decision once and records it, AI systems create and change their own algorithms to make decisions thousands, even millions of times, refining the results over time.[33] In some systems, AI may even review previous decisions using newly tuned algorithms to perform a self-audit and reveal potential issues.[34] The beauty of the loop is its ability to evolve and

---

perceive in that moment. *See* Diana L. DeMott & Roger L. Boyer, *And the Human Saves the Day or Maybe They Ruin It, The Importance of Humans in the Loop*, AMER. INST. AERONAUTICS & ASTRONAUTICS (2017), https://ntrs.nasa.gov/api/citations/20160013680/downloads/20160013680.pdf [https://perma.cc/T25W-LLZR].

[31] This Article does not aim to determine how liability might be allocated in AI cases, but rather identifies key liability issues.

[32] Crootof et al., *supra* note 1, at 440.

[33] *Id.*

[34] *See* McKendrick & Thurai, *supra* note 8.

refine. AI may be initially designed to function *without* a human in the loop.[35]

At its very foundation, AI is a closed system of automated decision-making, which includes at least one decisional "loop."[36] A closed system is not open to general inspection; it is operated in a proprietary manner consistent with principles of confidentiality and secrecy.[37] Even when AI systems continuously learn, their technical functions are protected from public disclosure.[38]

As explained by Rebecca Crootof, Margot Kaminski, and W. Nicholson Price II, a "human in the loop" may be defined as "an individual who is involved in a single, particular decision made in conjunction with an algorithm."[39] This working definition is helpful because there are several ways in which a human is involved in a decision in conjunction with an algorithm. The key differentiator is that a human is involved, at least once, *during* the moment a decision is being made.[40] Humans in the loop are expected to make decisions about system function, usually without knowing the range of decisions AI can make or the ways it can malfunction. Even when a human is supposed to intervene in an AI's function, the human cannot do so knowledgeably. To be defined as "humans in the loop," they must be involved in the decision in some way, even if passively, rather than reflectively evaluating after the fact (i.e., after the loop), as would be the case in AI auditing or

---

[35] Although humans are often placed in the loop, AI is usually designed without or for its human user, with designers focusing primarily on functionality with respect to the AI system's architecture and training data. Any design for humans, if human-centric design does occur, occurs after AI is already created. *See, e.g.*, Claire M. Felmingham et al., *The Importance of Incorporating Human Factors in the Design and Implementation of Artificial Intelligence for Skin Cancer Diagnosis in the Real World*, 22 AM. J. CLINICAL DERMATOLOGY 233 (2021) (describing how human factors may affect and influence clinical AI use and why integrating human factors knowledge into the design process can avoid these challenges).

[36] Crootof et al., *supra* note 1, at 440.

[37] Anees Ahmed, *Demystifying Closed Systems in Generative AI: Benefits, Challenges, and the Future*, AISTACKED (2023), https://aistacked.com/demystifying-closed-systems-in-generative-ai-benefits-challenges-and-the-future [https://perma.cc/JT54-37QC].

[38] *Id.*

[39] Crootof et al., *supra* note 1, at 440.

[40] *Id.* at 441-43.

other decision reviews. For this reason, humans reviewing content-monitoring decisions after the fact are not humans in the loop, but a human supervising an AI-enabled nuclear reactor would be.

A human in the loop can take different forms, from high-touch, collaborative work *with* AI to largely supervisory observation of the AI without direct intervention or joint action.[41] AI can operate intangibly, or virtually, to support tasks or to make recommendations. It can also direct physical functionality both within a device that connects to the Internet, the Internet of Things,[42] or outside a device or machinery through a wireless or wired connection from another computer. For example, AI can assess your interview video for "job fit" or generate a credit score.[43] AI can optimize machinery function or other processes, such as by delivering insulin through an insulin pump to maintain a normal blood sugar range or by managing traffic light timing to improve traffic flow.[44] The functional reason we place a human in the loop in

---

[41] *Id.* at 473. *See* Katie D. Evans, Scott A. Robbins & Joanna J. Bryson, *Do We Collaborate With What We Design?*, TOPICS COGNITIVE SCI., Aug. 15, 2023, at 1, 5-6 (questioning whether collaboration is really possible between humans and machines).

[42] Internet of Things devices have their main decisional software within a physical device, but this software can be updated regularly and is often designed to be "online," exchanging information over a wireless (Wi-Fi) connection. Charlotte A. Tschider, *Regulating the IoT: Discrimination, Privacy, and Cybersecurity in the Artificial Intelligence Age*, 96 DENV. L. REV. 87, 92 (2018) (describing the key parts of IoT as physical components, smart components, and connectivity components).

[43] Klara Cervenanska, *How Can AI Help Job Seekers? 5 AI Tools to Use in Your Next Job Search*, KICKRESUME BLOG (Feb. 20, 2024), https://blog.kickresume.com/how-can-ai-help-job-seekers [https://perma.cc/BFZ8-DZNF]; Cem Dilmegani, *AI Credit Scoring Models in 2024: In-depth Guide*, AIMULTIPLE (Dec. 23, 2022), https://research.aimultiple.com/credit-scoring-model [https://perma.cc/LM76-WN2Y].

[44] Zhouyu Guan, et al., *Artificial Intelligence in Diabetes Management: Advancements, Opportunities, and Challenges*, 4 CELL REPS. MED., Oct. 17, 2023, at 1, 6; Sagi Or, *AI in Transportation 2023: The Big Changes Ahead*, PTV GROUP (Feb. 15, 2023), https://blog.ptvgroup.com/en/real-time-

any of these scenarios is to reduce the inherent risk of using AI we don't trust.[45] However, AI creators may also exploit this functional purpose to avoid legal liability.

In the Sections that follow, I explain how AI's complexity limits the effectiveness of any human designated to supervise, intervene, or collaborate with an AI system. Although humans in the loop are not equipped with the knowledge to challenge an AI system, human decision-making in AI design significantly affects AI system safety, efficacy, and fairness. As explored in Parts II and III, this reality suggests that regulatory and liability models reinforcing the importance of human decision-making in the *creation* of AI systems will contribute more significantly to those goals than relying upon humans in the loop.

A. Deep Learning Neural Networks Are Complex and Inscrutable

The design of AI systems makes the role of a human in the loop largely symbolic. On the surface, humans might appear to be the perfect solution for a lack of confidence in AI: having a human in the loop simply makes us feel better about using it.[46] AI systems have highly complex structures, largely because complexity often improves output accuracy.[47] However,

---

traffic-management/ai-in-transportation-2023-the-big-changes-ahead [https://perma.cc/3SF9-EBQH].

[45] Juliette Powell, *Automation Complacency: How to Put Humans Back in the Loop*, UNITE.AI (Aug. 31, 2023), https://www.unite.ai/automation-complacency-how-to-put-humans-back-in-the-loop [https://perma.cc/2ZXN-H4J9]; Adrian Bridgwater, *When AI Needs a Human-In-The-Loop*, FORBES (June 2, 2023), https://www.forbes.com/sites/adrianbridgwater/2023/06/02/when-ai-needs-a-human-in-the-loop/?sh=7238f7d31b41 [https://perma.cc/2CB2-C6PG].

[46] Vijay Kotu, *Are You in The AI Loop?*, FORBES (Dec. 17, 2021), https://www.forbes.com/sites/servicenow/2021/12/17/are-you-in-the-ai-loop [https://perma.cc/9AY2-JD3J]; Bjorn Reynolds, *Remember the Human Element of the Emerging AI World*, FORBES (Aug. 30, 2023), https://www.forbes.com/sites/forbeshumanresourcescouncil/2023/08/30/remember-the-human-element-of-the-emerging-ai-world/?sh=19b7206e7f23 [https://perma.cc/J2HT-4ZG7].

[47] Accuracy, also called efficacy in medical device performance, is a key concern for a variety of AI models. Kathy McGrath, *Accuracy and Explainability in Artificial Intelligence: Unpacking the Terms*, FORTY-

complexity can lead to inscrutability, making it nearly impossible for a human to understand how an AI system made a decision that resulted in a given output. Any explanations of the decision would be approximations of complex algorithmic outputs and a flattened version of the actual AI decision.[48]

An article drafted by NASA risk analysts illustrates this point in its title: *And the Human Saves the Day or Maybe They Ruin It, The Importance of Humans in the Loop*.[49] Humans can potentially work with AI, but humans require special knowledge and training to effectively identify when AI is functioning improperly, limiting the universe of people who can safely supervise AI. In some cases, humans lack the necessary information or training to effectively challenge AI decisions because the AI is too complex to understand.[50] In other cases, AI may not disclose information because it could compromise confidentiality, which may be crucial for the business practices using AI. Disclosing information to consumers, for example, would mean that the information would also be readily available to competitors. Yet, with limited disclosures, AI users will have less information and will be less able to understand AI decisions.

SECOND INT'L CONF. INFO. SYS. 1, 2 (2021), https://bura.brunel.ac.uk/bitstream/2438/26392/4/FullText.pdf [https://perma.cc/SM8G-49YZ].

[48] For this reason, researchers have had to develop approximations for explanation that fit the goals of these systems. For example, the National Institute for Standards and Technology drafted four principles of explainable systems that, in part, describe how to communicate explanations that could be useful. *See* Jonathon P. Phillips et al., *Four Principles of Explainable Artificial Intelligence*, NISTIR 8312 (Sept. 2021), https://nvlpubs.nist.gov/nistpubs/ir/2021/NIST.IR.8312.pdf [https://perma.cc/U9SC-M2HE].

[49] *See* DeMott & Boyer, *supra* note 30, at 1.

[50] Michael Weber et al., *Organizational Capabilities for AI Implementations—Coping with Inscrutability and Data Dependency in AI*, 25 INFO. SYS. FRONTIERS 1549, 1558 (2023).

Scholars have discussed these concepts broadly under the umbrella of *opacity*.[51] Opacity is a general term that covers any mechanism for preventing information transparency and disclosure. Opacity could include technology, business, or legal practices that create barriers to information transparency.[52] The "black box" loop is both unavoidably opaque and intentionally opaque.[53] The loop functions with complex, inscrutable systems, which are unavoidably opaque.[54] Businesses may intentionally make the loop opaque to preserve the company's competitive advantage in the marketplace.

It is easy to raise concerns about opacity—after all, transparency is increasingly a goal individuals value. AI, however, may be worth its opacity cost. In structure, speed, connectivity, updatability, and scalability, computers beat out humans.[55] Humans do not have equivalent cognitive capacity in our physical brains,[56] while the most powerful computing systems have almost limitless capacity. When it comes to AI, at least for certain types of reasoning, "we probably are not as smart as we think."[57] Indeed, AI is designed not to replicate existing human processes but to transform them.[58] Yet, AI that addresses complex problems will often be inscrutable to

---

[51] Opacity, as many scholars have adopted this term, is something that can be overcome through forcing explanation. *See, e.g.*, Warren J. von Eschenbach, *Transparency and the Black Box Problem: Why We Do Not Trust AI*, 31 PHIL. & TECH. 1607, 1611 (2021). Practically speaking, however, the inscrutable nature of technology is not something that can be overcome easily. *Id.* at 1608.

[52] FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 6-7 (2015); W. Nicholson Price II, *Black-box Medicine*, 28 HARV. J.L. & TECH. 419, 433-34 (2015); Charlotte A. Tschider, *Legal Opacity: Artificial Intelligence's Sticky Wicket*, 106 IOWA L. REV. 126, 131-34 (2021).

[53] *See* Tschider, *supra* note 52, at 129.

[54] Charlotte A. Tschider, *Beyond the "Black Box*," 98 DENV. L. REV. 684, 690-91 (2021).

[55] *See* Giulio Coppi, Rebeca Moreno Jimenez & Sofia Kyriazi, *Explicability of Humanitarian AI: A Matter of Principles*, 6 J. INT'L. HUMANITARIAN ACTION, Oct. 6, 2021, at 1, 5.

[56] *Id.* at 3.

[57] *Id*.

[58] *See* Thomas M. Powers & Jean-Gabriel Ganascia, *The Ethics of the Ethics of AI*, *in* THE OXFORD HANDBOOK OF ETHICS OF AI 26, 49 (2020).

humans.[59] Such limitations may present a problem if society's goal is to impose shared responsibility and shared, shifted, or transferred liability.

Machine learning is the primary type of AI people interact with today and is the foundation for deep learning applications powered by neural networks.[60] As discussed *supra*, these complex applications can produce results that are unexplainable or uninterpretable by humans, rendering the AI's decision-making unintelligible. Machine learning leverages tremendous computing power to identify relationships between substantial volumes of data elements, usually relationships a human, even an expert, would not be able to identify on their own.[61] It encompasses a wide variety of approaches and techniques, including neural networking (specifically, deep learning) and natural language processing,

---

[59] *See* Tschider, *supra* note 54, at 722 (arguing that although explanations could be necessary in some situations, model outputs through testing will be far more important for building safe and fair AI systems); Lilian Edwards & Michael Veale, *Enslaving the Algorithm: From a "Right to an Explanation" to a "Right to Better Decisions"?*, INST. ELEC. & ELECS. ENG'RS SEC. & PRIV., May-June 2018, at 1, 46; Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1111 (2018) (describing a conventional mindset that inscrutability increases accuracy and reduces the potential for explainable AI). *But see* Joanna Bryson, *The Artificial Intelligence of the Ethics of Artificial Intelligence*, *in* THE OXFORD HANDBOOK OF ETHICS OF AI 2, 8-9 (2020) (describing the use of low-complexity models and audit frameworks to explain complex models, also suggesting that highly opaque systems could be analyzed using black-box forensic methods).

[60] *Machine Learning vs. AI: Differences, Uses, and Benefits*, COURSERA (June 16, 2023), https://www.coursera.org/articles/machine-learning-vs-ai [https://perma.cc/2E34-6GZ5].

[61] Human intuition can overcome computing performance in some cases, but the exact path to such intuition is largely unexplainable, making the process difficult to simulate. Christopher Larkin, *AI Won't Replace Human Intuition*, FORBES (Sept. 27, 2022), https://www.forbes.com/sites/forbestechcouncil/2022/09/27/ai-wont-replace-human-intuition/?sh=71eaea2767bf [https://perma.cc/D7W7-NHLE] (describing how AI cannot directly replace many human jobs).

aimed at enabling computers to learn from data and perform tasks without being explicitly programmed to do so.[62]

Neural networks are a distinct type of machine learning that use several computational layers to provide an output.[63] Neural networks are hallmarked by their ability to provide a high degree of predictive confidence on complex problems, such as a predictive tool analyzing hundreds or thousands of data inputs and providing, for example, the likelihood of a patient having a disease like Alzheimer's as an output.[64] Each computational layer uses machine learning, where a layer's output is used in the next layer's computation.[65] These layers also include probabilities, where the outcome for each is given a specific weighting. The final output of hundreds or thousands of layers of algorithmic computation provides the recommendation or automated direction.[66] The more layers used, the "deeper" (and more complex) the neural network becomes.[67] Figure 1 below provides a visual representation of a neural network that results in a shopping recommendation. The deeper the neural network, the more difficult it becomes to explain, and the more difficult it is to adequately communicate its function to a human in the loop.[68]

---

[62] *What is Natural Language Processing (NLP)?*, IBM, https://www.ibm.com/topics/natural-language-processing [https://perma.cc/NFH3-RS66].

[63] *See* Tschider, *supra* note 54, at 691-92.

[64] David Ahmedt-Aristizabal et al., *Graph-Based Deep Learning for Medical Diagnosis and Analysis: Past, Present and Future*, ARXIV 16-17 (May 27, 2021), https://arxiv.org/pdf/2105.13137.pdf [https://perma.cc/Y9HF-MWTV] (describing Alzheimers diagnosis as a potential use case for medical AI and machine learning).

[65] Tschider, *supra* note 54, at 691; Samuel K. Moore, David Schneider & Eliza Strickland, *How Deep Learning Works > Inside the Neural Networks That Power Today's AI*, IEEE SPECTRUM (Sept. 28, 2021), https://spectrum.ieee.org/what-is-deep-learning [https://perma.cc/DDU2-UE6F].

[66] Tschider, *supra* note 54, at 692.

[67] *Id.* at 691.

[68] Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai [https://perma.cc/L4SC-JVBF].

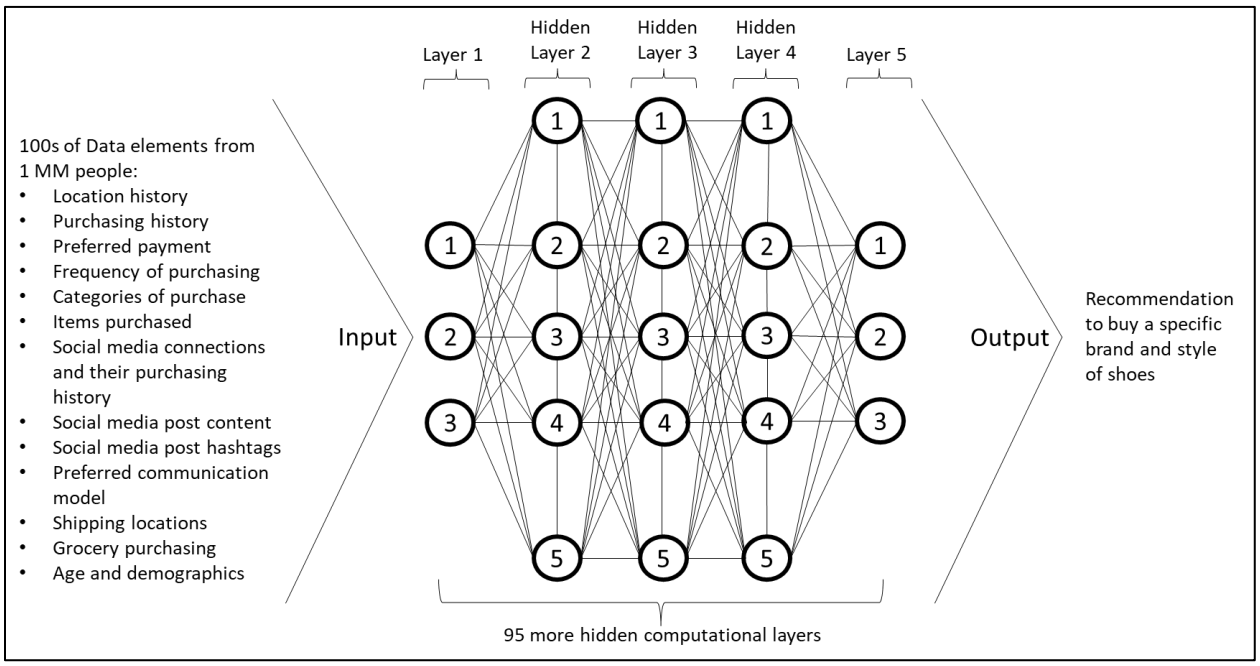

Figure 1: Example of a Deep-Learning Neural Network

Unfortunately, thousands of computational layers are often needed to refine which (and to what extent) data inputs translate to probabilities that these models are designed to output.[69] Many deep-learning models focus on solving human problems that are otherwise nearly impossible or would take countless hours to solve, such as how to diagnose cancer prior to clinically relevant symptoms presenting themselves, or how to avoid accidents in self-driving cars.[70] In both cases, the available data that could potentially be useful in achieving

[69] Gao Huang, *Deep Networks and Stochastic Depth*, *in* COMPUTING VISION—ECCV 2016, at 646, 646 (2016).

[70] Benjamin Hunter, Sumeet Hindocha & Richard W. Lee, *The Role of Artificial Intelligence in Early Cancer Diagnosis*, 14 CANCERS 1, 6 (2022) (describing the need for multi-variable data to diagnose cancer earlier); Rahib Abiyev & Murat Arslan, *Vehicle Detection Systems for Intelligent Driving Using Deep Convolutional Neural Networks*, 3 DISCOVER A.I., May 2, 2023, at *1, *1-2.

these goals are widely variable and high volume, requiring additional layers to process and interpret them.

In short, the added refinements of complex AI through these computational layers will likely be safer, fairer, and more effective than a simplistic version—presuming that the AI creator has developed AI effective for its intended use and architected the solution well. Conversely, the more complex an AI system is, the more likely its functions will be unintelligible to a human in the loop. The same design decisions that make an AI system function better may make it fundamentally more difficult to oversee effectively.

B. Generative AI Creates Similar Complexity Challenges

Another type of machine learning increasingly used in AI products is Generative AI (GAI). GAI models rely on machine learning algorithms to learn the underlying patterns and structures in training data so that they can generate new data samples that are similar.[71] For some GAI systems, the data supplied for training include linguistic expression, though GAI systems may also be trained to output images, video, or audio. Notably, large language models (LLMs) serve as the underlying language generation component for GAI.[72] When users interact with a GAI system, such as a content generator or chatbot, the system often utilized a pre-trained LLM to understand user input, generate responses, or create content based on the input and its internal model of language.

Although GAI might seem harmlessly expressive, they are positioned to power chat and other communication-based tools that interact with humans and direct human behavior.[73] Like

---

[71] *See* Huang, *supra* note 69, at 646.

[72] Ryan O'Conner, *Introduction to Large Language Models for Generative AI*, ASSEMBLYAI (May 17, 2023), https://www.assemblyai.com/blog/introduction-large-language-models-generative-ai [https://perma.cc/HH6Y-UWFN]. GAI relies on LLMs for interaction due to LLMs' advanced language understanding and generation capabilities. LLMs are trained on massive datasets containing a range of text, enabling them to understand, context, semantics, and grammar, and even generate coherent and contextually relevant responses. *Id.*

[73] Christopher Graves, *Generative AI Can Help You Tailor Messaging to Specific Audiences*, HARV. BUS. REV. (Feb. 16, 2023), https://hbr.org/2023/02/generative-ai-can-help-you-tailor-messaging-to-

other complex AI, GAI are largely unintelligible but generally gain effectiveness through their complexity. This means that a human could be harmed when they rely on GAI outputs, for example being provided the wrong medical diagnosis, but a human may not be able to interrogate and correct the system effectively.

C. AI Is More Than an Algorithm

To really understand how and to what extent humans fit into the loop, it is important to understand how humans are currently involved in the creation of AI. Data scientists, engineers, architects, database architects, security professionals, and any number of humans create AI systems. The involvement of humans *outside* the loop who create the loop and evaluate its function are crucial to the development of quality AI. Ultimately, through their design choices, these professionals have the potential to create safe, efficacious, and fair AI.[74] But if these professionals make poor design choices, then human intervention *inside* the loop is not likely to remedy the harmful impacts of the decisions of those outside the loop.

---

specific-audiences; Paul Smith, *AWS HealthScribe Uses Generative AI and Real-time Doctor-Patient Conversations to Update Medical Charts*, FORBES (July 28, 2023), https://www.forbes.com/sites/moorinsights/2023/07/28/aws-healthscribe-uses-generative-ai-and-real-time-doctor-patient-conversations-to-update-medical-charts [https://perma.cc/5D26-6BYY].

[74] Throughout this Article, I reference "safe, efficacious, and fair" AI to encompass the potential harms that could arise. Reasonably safe products balance risk and benefit in such a way that reduces risk to people and property. Efficacious products, such as those reviewed by the FDA, work or work for their intended purpose. Fair products are those that provide reasonably consistent service and options between individual people. Specifically, unfair products may be unsafe to specific individuals but safe to others. They may actually be discriminatory in that they are safe for many but unsafe for a protected class of individuals. Unsafe, non-efficacious, and unfair products will likely give rise to liability. The goal explained in this Article is to reduce the likelihood of harm while also providing a means of compensation in the event remaining individuals or organizations are harmed. Sharona Hoffman, *Medical Big Data and Big Data Quality Problems*, 21 CONN. INS. L.J. 289, 295-98 (2014); Charlotte A. Tschider, *Prescribing Exploitation*, 82 MD. L. REV. 857, 885 (2023).

For this reason, liability schemes should not treat the presence of a human in the loop as insulating those outside of it.

Human AI developers make design choices to determine what AI model and approach will be used, how AI will be tested and tuned, and how feedback will be integrated.[75] These design choices are crucial: they are the foundation for how AI will function and what problems (and potential liability) the AI could have. Preventing these problems at the design stage can mean avoiding property damage, protecting humans, avoiding using copyright-protected material, preventing hallucinations,[76] or limiting any number of other problems. Importantly, design choices need not be set in stone. For example, many developers are interested in creating AI systems that learn (and change) in real-time. These systems are designed to be "dynamic," or unlocked, meaning that they have the capacity to learn and change continuously, allowing users to see these changes immediately.[77] Even "locked" AI will usually have a new version running behind the scenes, gradually being revised as the data provided to the public AI change.[78]

Data scientists and engineers perform training, testing, and tuning activities that (ideally) ensure AI systems are performing as expected. When engineers train AI, they feed information to the AI to help it achieve the goal for which the

---

[75] For example, bigger data is not necessarily the ingredient for AI to meet its goals; while big data is essential to most complex AI, the nature of that data (including volume, type, diversity) is highly contextual. Bigger is not always better. *See* Bryson, *supra* note 59, at 17.

[76] Tim Keary, *AI Hallucination*, TECHOPEDIA (Jan. 15, 2024), https://www.techopedia.com/definition/ai-hallucination [https://perma.cc/YP5F-GFSP].

[77] Ben Lorica, *Why Continuous Learning Is Key to AI*, O'REILLY (Aug. 7, 2017), https://www.oreilly.com/radar/why-continuous-learning-is-key-to-ai [https://perma.cc/8JKQ-XX4D].

[78] Organizations may choose to release new versions of AI on a timeline by running both a production version (that has previously learned) and a secondary version that continues to learn until it can be tested enough to be released to production. Charlotte A. Tschider & Cynthia Ho, *Artificial Intelligence and Intellectual Property in Healthcare Technologies*, *in* RESEARCH HANDBOOK ON HEALTH, AI, AND THE LAW (Barry Solaiman & I. Glenn Cohen eds.) (forthcoming 2024) (manuscript at 8 n.46) (on file with authors).

AI is designed,[79] for example predicting recidivism rates.[80] While the initial design determines how the AI is structured and learns, training is what allows the AI to understand what outputs it should generate when given certain inputs.[81]

When training is not designed effectively, for example, if humans train the algorithm using low-volume, highly similar, non-contextual, or non-representative data, the trained algorithm is likely to produce unsafe or discriminatory outputs.[82] Algorithms may make approximations for data they do not have or develop inferences based on the non-representative data on which they have been trained. AI without an effective training environment or appropriate data for its purposes creates greater risk for AI use—risk that a human in the loop cannot overcome.[83]

---

[79] Brian Bailey, *Defining and Improving AI Performance*, SEMICONDUCTOR ENG'G (Dec. 12, 2019), https://semiengineering.com/defining-ai-performance [https://perma.cc/ATE7-GZYD] (describing performance characteristics not just in terms of accuracy but also energy consumption); Will Douglas Heaven, *The Way We Train AI Is Fundamentally Flawed*, MIT TECH. REV. (Nov. 18, 2020), https://www.technologyreview.com/2020/11/18/1012234/training-machine-learning-broken-real-world-heath-nlp-computer-vision [https://perma.cc/5DWX-3R7Y] (identifying key problems in application from training environments to real-world applications).

[80] Mirilla Zhu, *An Algorithmic Jury: Using Artificial Intelligence to Predict Recidivism Rates*, YALE SCI. (May 15, 2020), https://www.yalescientific.org/2020/05/an-algorithmic-jury-using-artificial-intelligence-to-predict-recidivism-rates [https://perma.cc/B4JS-JSJA].

[81] Shehmir Javaid, *5 AI Training Steps & Best Practices in 2023*, AIMULTIPLE (Aug. 28, 2023), https://research.aimultiple.com/ai-training [https://perma.cc/422U-G6B6] (describing the importance of training and validation of training outcomes).

[82] *See* W. Nicholson Price II, *Medical AI & Contextual Bias*, 33 HARV. J.L. & TECH. 66, 99 (2019) (explaining how AI exacerbates existing issues of contextual differences in healthcare); *see* Tschider, *supra* note 74, at 885-86.

[83] Assumption of risk in these situations also creates challenges, as for complex systems explaining potential risks requires that such risks are foreseeable and can be communicated with some degree of detail for licensees or users to provide their informed consent. *See* Wex, *Foreseeable Risk*, CORNELL L. SCH. (Dec. 2022), https://www.law.cornell.edu/wex/foreseeable_risk

When humans validate AI, they provide a "trained" AI with non-training data to uncover potential issues,[84] including biases, safety issues, and potential inaccuracies.[85] Developers can validate and test AI to reveal issues before AI is made public, avoiding any number of more serious problems that could harm people, organizations, or property once the AI is publicly released. When developers create a testing environment that does not adequately reflect the choices the AI will need to make, the prompts it will receive, or the real-world data it will experience or perceive, the tests cannot lead to quality AI.[86] In a world where AI could perform in unexpected ways due to its opacity, including lying or hallucinating, the public cannot afford to inherit issues that are reasonably preventable through effective testing.[87]

Model and data selection, training, testing, and validation are essential steps to develop safe, efficacious, and fair AI. However, otherwise quality algorithmic design can be thwarted if developers apply poor system and infrastructure design. System technology components and infrastructure establish affordances and constraints for a system, affecting security and performance that in turn shape the overall safety, efficacy, and

---

[https://perma.cc/U4WH-J3K5]. It could also be noted that use of AI likely will affect individuals who would not be able to accept risk, for example, a factory operator killed by AI-directed machinery or a resident of a county denied access to housing.

[84] *See* Javaid, *supra* note 81.

[85] *Id.*; Apostol Vassilev, Harold Booth & Murugiah Souppaya., *Mitigating AI/ML Bias in Context: Establishing Practices for Testing, Evaluation, Verification, and Validation of AI Systems*, NIST NAT'L CYBERSECURITY CTR. EXCELLENCE (Nov. 2022), https://www.nccoe.nist.gov/sites/default/files/2022-11/ai-bias-pd-final.pdf [https://perma.cc/8SLH-KVA2] (laying out critical considerations for testing practices to mitigate bias, resulting in a future NIST publication).

[86] *See* Javaid, *supra* note 81. For example, AI that has not been subjected to Algorithmic Impact Assessments or similar testing protocols to avoid discriminatory impact will not result in quality AI.

[87] Catherine Thorbecke, *AI Tools Make Things up a Lot, and That's a Problem*, CNN BUS. (Aug. 29, 2023), https://www.cnn.com/2023/08/29/tech/ai-chatbot-hallucinations/index.html [https://perma.cc/LU94-NRW9]. While GAI can certainly create facts to sound plausible, other AI can fill in essential gaps to make decisions. In either case, the potential issues resulting from inadequate training and validation (and poor data selection) can be numerous.

fairness of the AI product.[88] Data scientists work with engineers and system architects to design systems that can support the desired performance of a given AI model, but if those developers design the infrastructure poorly, they can prevent the model from effectively reaching its codified goals or achieving a desired level of accuracy.[89] For example, AI requires CPU or GPU and RAM (characteristics of physical computers) with much greater capacity than other types of technology.[90] Whether or not AI has enough processing or storage capacity determines what programming an AI system could need, including workarounds that could compromise more straightforward designs or introduce performance or security risks.[91]

Developers must plan for critical capabilities to maintain safe, accurate, and fair AI. For example, business continuity and cybersecurity may be viewed as ancillary to AI development, but these capabilities may affect AI design, depending on the AI's function.[92] If AI manages a system of locks that control water release to rivers and streams, a ransomware attack that prevents the system from releasing

---

[88] *See* Tigran Hovsepyan, *A Guide to Artificial Intelligence Infrastructure*, PLATAI (Oct. 28, 2022), https://plat.ai/blog/artificial-intelligence-infrastructure-guide [https://perma.cc/DY4F-XUQA].

[89] *See* Bryson, *supra* note 59, at 6-7.

[90] James Montantes, *Hardware Requirements for Artificial Intelligence*, MEDIUM (May 14, 2021), https://becominghuman.ai/hardware-requirements-for-artificial-intelligence-653335df899f [https://perma.cc/W2DZ-GVYP].

[91] *Id.*

[92] *See* Tschider, *supra* note 42, at 133. *See generally* Andrea M. Matwyshyn, *The Internet of Bodies*, 61 W&M L. REV. 77, 83 (2019) (describing a 2017 Internet-of-things failure causing an interruption in a surgery); Charlotte Tschider, *Deus ex Machina: Managing Cybersecurity Threats for Patients of the Future*, 5 SAVANNAH L. REV. 177, 190-91 (2018) (examining the absence of effective FDA regulation relating to cybersecurity threats affecting patient safety in AI and recommending regulatory strategies); Scott J. Shackelford et al., *Securing the Internet of Healthcare*, 19 MINN. J.L. SCI. & TECH. 405, 414-16 (2018) (describing the use of blockchain technology to secure medical device technologies).

water could result in severe flooding upstream.[93] If AI technology simply stops working in the middle of a laparoscopic microsurgery using a surgical robot, it is possible a patient could experience severe injury, even if a doctor is operating the tools.[94] So the selection and design of overall system components and infrastructure, including business infrastructure, can have a serious impact on AI's functionality.

In addition to the system and infrastructure, the choices developers make with respect to data selection, storage, and architecture can seriously affect the safety, efficacy, and fairness of AI systems. After all, machine learning applications rely principally on deep, accurate, contextually rich, and representative data. Unlike typical software systems, AI requires substantially large databases on servers, which may require greater data storage capacity. These servers may be located in different geographic locations if required to comply with data protection or data localization requirements.[95] AI systems also have interactive components like the user interface in a Web application or software that caches data for faster performance.[96] All servers are connected within one

---

[93] Charlotte A. Tschider, *Locking Down "Reasonable" Cybersecurity Duty*, 41 YALE L. & POL'Y REV. 75, 78 (2023). Ransomware attacks, which deny use of some resource until a ransom is paid to a hacker, have caused major disruptions in critical infrastructure, including, for example the Colonial Pipeline attack, which impacted the transportation of gasoline and other refined oil products to the East Coast in 2021. *See* Congressional Research Service, *Colonial Pipeline: The DarkSide Strikes* IN11667, CRS INSIGHT (May 11, 2021), https://crsreports.congress.gov/product/pdf/IN/IN11667 [https://perma.cc/C6QR-C3P6].

[94] The availability of certain kinds of data, including those used to inform the precise moves of surgical robots, are essential to safety in a variety of situations. *See* Matwyshyn, *supra* note 92, at 83. This is why HIPAA requires under the Security Rule the development of emergency plans related to protected health information. HHS Data Security Standards, 45 C.F.R. §§ 164.306-.318 (2022). While few other statutes require this, it illustrates how AI could require additional planning for service disruptions.

[95] Data protection is the terminology used to include both privacy and cybersecurity considerations outside of the United States.

[96] A user interface is the layer of a technology a human person using the AI interacts, such as a mobile app or web application. Cache is the part of a computing system that retains data temporarily. Josh Lovejoy, *The UX of AI*, GOOGLE DESIGN, https://design.google/library/ux-ai [https://perma.cc/7K5X-ZAH9]; Piyush, *Intelligent Caching: Machine*

hardware unit, via intranet (i.e., an internal protected network), or over the Internet (i.e., a public and open network), using a variety of different security architectures and strategies for protecting the AI from unauthorized access. Humans are involved in every one of these decisions, from hardware to software, to the use of third-party service providers and products in a "technology stack" powering AI, to the overall system architecture.[97] Despite society's focus on the algorithms and user interface as AI, system and architecture choices also affect the safety and security of AI models.

The various third-party entities that play a role in the overall function of the AI further complicate the picture of AI design. Although we have explored AI as the creation of one organization, modern AI is just like any manufactured product, created from a variety of parts. Each of the discretionary design choices described in this Section may include third parties and third-party tools, physical and virtual components, software, computing environments, and consultants.[98] A "technology stack," or the selection of technology used collectively in one product, like an autonomous vehicle or a diagnostic cancer app, also incorporates potential risks that other technologies in the stack introduce.[99] The AI stack functions holistically, but potential issues with various parts of the stack (such as cybersecurity vulnerabilities) could radically change the safety of the whole system. Overall, the relative safety, efficacy, and fairness of a system results from many distinct choices, some of which may be made by various organizations.

While organizations should generally consider all of the above factors prior to releasing AI to the public or licensing or

---

Learning and AI in Memory Optimization, PE BLOGS (July 14, 2023), https://perfectelearning.com/blog/intelligent-caching-machine-learning-and-ai-in-memory-optimization [https://perma.cc/Z6NZ-KNL4].

[97] *See* Bryson, *supra* note 5959, at 23.

[98] Ayush Patel, *The Complete Guide to the Modern AI Stack*, TOWARDS DATA SCI. (June 4, 2022), https://towardsdatascience.com/the-complete-guide-to-the-modern-ai-stack-9fe3143d58ff [https://perma.cc/8BRY-5BWU].

[99] *Id.*

selling AI to another organization, it is important to note that AI design, training, and testing is informed by how well the AI model works when used.[100] The "loop," therefore, can be designed for further change and refinement in future loops—feedback loops.[101] It is continuous, not static, and should result in safer, more accurate, and fairer processes over time.[102] Ultimately, safe, efficacious, and fair AI must be responsibly developed and responsibly managed over its lifetime. Although including a human in the loop might make us feel safer, unsafe AI cannot be reliably and routinely saved by a human in the loop—it can only be saved by better design. Reinforcing the idea of a human in the loop as safeguarding AI obscures underlying problems. If we knew a human in the loop was not available, we would likely expect better design from AI creators. More importantly, if, given the complexities of AI design, the AI creator does not even know of all the potential

---

[100] Although pre-release testing is crucially important, nothing can prepare for broad, extensive, diverse use of an AI system to identify issues. Theresa Smith, *Test Your AI for the Real World*, SPHEREOI (Sept. 2, 2022), https://sphereoi.ai/test-your-ai-for-the-real-world [https://perma.cc/8CWS-CGDD].

[101] Feedback loops are used to refine AI algorithms for purposes of better accuracy or performance. Gesche Loft, *How AI Uses Feedback Loops to Learn From Its Mistakes*, ULTIMATE (Sept. 4, 2023), https://www.ultimate.ai/blog/ai-automation/what-is-a-feedback-loop [https://perma.cc/V7N7-D3SL]. However, different feedback loop designs may have very different results. Nicolò Pagan et al., *A Classification of Feedback Loops and Their Relation to Biases in Automated Decision-Making Systems*, 9TH INT'L CONF. COMPUTATIONAL SOC. SCI. IC^{2}S^2 at 2 (July 17-20, 2023), https://nicolo-pagan.github.io/research/abstracts/IC2S2__Feedback_Loops.pdf [https://perma.cc/WHP4-WJY7] (describing the various versions of feedback loops and their effect on prediction models and potential for bias). AI live testing methods often restrict continuous learning or use old training data. Newer methods use a live instance to refine AI characteristics and improve performance. Jaya Krishna Mandivarapu, Blake Camp & Rolando Estrada, *Self-Net: Lifelong Learning via Continual Self-Modeling*, 3 FRONTIERS A.I. 1, 2-3 (Apr. 9, 2020).

[102] Natalie Fletcher, *Closing the Loop: How Feedback Loops Help to Maintain Quality Long-Term AI Results*, CLARIFAI (May 24, 2019), https://www.clarifai.com/blog/closing-the-loop-how-feedback-loops-help-to-maintain-quality-long-term-ai-results [https://perma.cc/B5WB-CL3W]; Joseph Harisson, *What Is AI Feedback Loop?* IT COMPANIES (Feb. 19, 2024), https://itcompanies.net/blog/ai-feedback-loop [https://perma.cc/UHU5-EW24].

risks of a given AI model, it is dangerous to assume that a human in the loop will.

D. The Human in the Loop Cannot Overcome Poor Design

Since the very beginnings of technology, humans have examined their relationships with it. Whether designing better technology for human use, replacing humans with technology, or developing collaborative technologies to enhance or supplement human activities, human have technology's potential to benefit society—generally with a human involved. Countless technology, psychology, and social science disciplines have studied the relationship between humans and technology.[103]

Humans' ability to work with technology (and the technology needed to achieve this relationship) spurred the areas of human-computer interaction (HCI), usability, and user experience (UX), at least as early as Don Norman's 1988 work, *The Psychology of Everyday Things*.[104] These fields have evolved to extend to more dynamic relationships between humans and technology, referred to as Human-Computer Collaboration—a field combining HCI and AI.[105] The relationships between humans and technology are nothing new—many people have made a career out of designing technologies for humans or training humans to use technology.

A human in the loop, however, involves more than just a relationship between humans and technology. A proactive decision to place a human in the loop seems to suggest: "This is dangerous—a human could help" or perhaps "We aren't sure this works correctly—a human can ensure it does." Humans are found in any number of loops, frequently when a human's presence is mostly ceremonial and objectively ineffective—

---

[103] KENT L. NORMAN, CYBERPSYCHOLOGY: AN INTRODUCTION TO HUMAN-COMPUTER INTERACTION 7-10, 64, 104-07 (2008).

[104] DON NORMAN, THE DESIGN OF EVERYDAY THINGS xiv, xviii (2013). *See also* Brad A. Myers, *A Brief History of Human-Computer Interaction Technology*, 5 ACM INTERACTIONS 44, 46 (1998).

[105] Loren G. Terveen, *Overview of Human-Computer Collaboration*, 8 KNOWLEDGE-BASED SYS. 67, 68 (1995).

what Crootof, Kaminski, and Price dub "warm bodies."[106] Although a human's role in the loop sometimes truly involves collaboration with AI, more often than not the human is not intended to bring some special skill that is necessary to AI model's function. Because the loop itself is a closed system that involves complex and automated decision-making, a human in the loop may not be able to knowledgeably challenge or meaningfully intervene in automated decisions. Below, Table 1 illustrates AI decision-making types that typically involve a human in the loop. These decision-making types are organized from least human-in-the-loop involvement to most.[107]

Designed Decision Type	**Embodiment**	**Human Effort in the Loop**	**Description**	**Example**
Fully Autonomous	Virtual	None	AI that does not have physical function operates without any involvement of a human but is used to the benefit or detriment of a person.	A person applies for a car loan on a website that uses AI to make loan decisions. The AI makes the decision, including the amount for which the person is qualified, and the decision is non-reviewable.
Fully Autonomous	Physical	None	AI that has a physical output directed by an AI decision without any involvement of a human.	A wind turbine automatically adjusts its speed, including stopping, starting, or locking its blades based on current or future weather conditions,

[106] *See* Crootof et al., *supra* note 1, at 485.

[107] This table is organized by function of the human in the loop in relation to the AI. These AI types are specifically grouped for purposes of analyzing their respective tort systems and potential limitations. Crootof et al. offer a more detailed account of AI roles and their legal positioning. *See* Crootof et al., *supra* note 1, at 473-87.

Designed Decision Type	Embodiment	Human Effort in the Loop	Description	Example
				optimal electricity generation, and potential income generation versus turbine lifespan. The turbine's function may be inspected periodically but not during its autonomous function.
Supervision/ Monitoring	Usually physical	Low	Technology is supervised continuously by a human while it is running.	A construction worker supervises as AI-enabled robots construct the walls of a home. The worker sets up the machine, monitors it in operation, and realigns it when needed.
Intervention (often combined with Supervision)	Usually physical	Medium	A human only intervenes in decisional processes in emergency situations.	A driver of an autonomous semi-truck observes that the truck is driving into oncoming traffic, reading a two-lane highway as two lanes in the same direction. The driver overrides the system and moves the truck into the correct lane.

Designed Decision Type	Embodiment	Human Effort in the Loop	Description	Example
Collaboration[108]	Virtual or physical	High	Technology requires heavy inputs from a human with special skills to make any decision or recommendation.	A radiology technician operates a Magnetic Resonance Imaging (MRI) machine to capture images of a patient's lungs. The patient's doctor inputs other data, including blood panel details, comorbidities, and age. A radiologist uses an AI diagnostic tool to process the data. The radiologist reviews the AI-generated diagnosis and confirms it. The patient's doctor refers them to an oncologist, who uses another AI to design a course of treatment.

Table 1: AI Decisional Systems

The categories in Table 1 illustrate the degree to which humans may be expected to be "in the loop." In truly collaborative systems that are designed to optimize *both* human and computer involvement, a human may work with AI

[108] The rise of "collaborative" AI is significant; forty-three percent of respondents to a Deloitte survey reported that they have a leader responsible for human and AI collaboration. DELOITTE, STATE OF AI IN THE ENTERPRISE REPORT 17 (5th ed. Oct. 2022), https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-ai-institute-state-of-ai-fifth-edition.pdf [https://perma.cc/HD5Q-CNQR].

to get the most beneficial output.[109] Many functions of humans in the loop, such as supervisory functions, are comparatively passive even though humans are evaluating decisions in the loop. In those contexts, it is, logically, unlikely that a human in the loop will perfect an AI model's function simply by being there. In contrast, technologies that require humans to make decisions collaboratively should involve humans with special expertise, such as airplane pilots or surgeons.[110] Thus, relying too heavily on humans in the loop for non-collaborative relationships may obscure a focus on essential design decisions that directly impact AI safety, efficacy, and fairness. Further, reliance on humans in the loop may also obscure who is actually responsible when things go wrong.

II. Looped Liability

AI creators are legally protected by humans in the loop, whether those humans are downstream consumers using AI products for their benefit or people monitoring AI function as part of their job. Although different regulatory and liability legal approaches can apply to these contexts depending on the type of AI used, nearly all approaches offer advantages to organizations that use humans in the loop. Creators of AI systems may attempt to receive less comprehensive review *ex ante* (e.g., in high-touch regulatory environments).[111] Or, they

---

[109] As described in Part II, *infra*, however, while human-computer collaboration may occur in some respects reasonably effectively for purposes of satisfying a specific goal like "assemble a car," the concept should not be mistaken for equal footed power dynamics, information symmetry, or effective liability shifting.

[110] Although these individuals may not have sufficient information to perform this role well, due to the inscrutability of AI systems, they are likely better positioned than non-expert humans to challenge AI because they possess expertise that characteristically has not been dependent on AI.

[111] For example, medical device manufacturers may seek 510(k) clearance rather than a pre-market approval process for a substantially equivalent device that may not actually be substantially equivalent, due to its introduction of AI to a preexisting device. Or, an organization may include various disclaimers to an AI diagnostic app so that it is not regulated by the

may contractually shield themselves from liability or attempt to shift liability in tort actions *ex post*.[112] Preemptively, organizations creating products include limitations on liability or disclaimed implied warranties in contracts to limit potential risk when contracting with business or otherwise non-consumer entities. Indeed, the U.S. tort system as it exists today will permit AI manufacturers to operate with almost no risk.[113] The proceeding Sections examine the many ways in which existing liability models are not designed for AI, and how humans in the loop will likely shield the creators of unsafe, ineffective, or unfair AI from liability.

A. The AI Supply Chain Involves Various Actors

To understand the legal dynamics in AI systems, it is first important to understand how AI is created, used, licensed, and acquired. The relationships between entities can have a substantial bearing on the availability of legal remedies and the type of claims a plaintiff can bring.[114] As described in Part I, AI

FDA but only by the FTC. AI may also be classified separately, largely avoiding heavy *ex ante* review. Charlotte A. Tschider, *Medical Device Artificial Intelligence: The New Tort Frontier*, 46 BYU L. REV. 1551, 1576, 1611 (2021).

[112] This liability shield, at least for many sectors, does not involve preemption, except when used for sectors that enjoy field preemption or statutorily created preemption. *See, e.g.*, *id.* (describing tort preemption for medical devices under the MDA and the likelihood of tort recovery in negligence).

[113] There are a wide variety of liability models not discussed here, for example vicarious liability, joint liability, several liability, joint and several liability, joint enterprise liability, and any variety of impleaded third parties in these lawsuits. Many or most of these liability models will suffer from similar limitations to those discussed here. *See* Joseph Angland, *Joint and Several Liability, Contribution, and Claim Reduction*, 3 ISSUES COMPETITION L. & POL'Y 2369, 2371-76 (2008) (arguing that joint liability can lead to fundamental distributional fairness claims when plaintiffs can decide who pays); *see also* Agnieska McPeak, *Sharing Tort Liability in the New Sharing Economy*, 49 CONN. L. REV. 171, 195-97 (2016) (describing joint enterprise liability as one model of shared liability when people are engaged in a common business purpose). Joint enterprise liability arises when a duty is created by another law (it does not create an independent duty). *Id.* at 195 n.130.

[114] Although not explored in extensive detail in this Article, one of the main issues that can affect these cases is privity, whether privity in contract or

design choices, including which third-party services may be used, affect potential downstream liability. Organizations design AI products such as software, web applications, and mobile device apps; manufacturers may license or purchase these products and either sell them or use them in physical devices.[115]

AI development companies, acquirers of AI, distributors, licensees, and front-line companies interface with a variety of users, in various combinations. An AI-development company could have a direct relationship with a downstream user, or the AI development company could be upstream from multiple other companies. The nature of the directness of this relationship affects who will face potential liability when things go wrong. For example, an AI creator could license its product to a development company that uses the AI for operations software that is distributed by that company. The software is used by employers to manage projects in a manufacturing plant. An employee uses the operations software to instruct various machines to produce certain products. When the AI fails, who is liable? Is it the AI creator, the licensor that integrated the AI into a new product, the employer, or the employee who was tasked with overseeing the AI-enabled software? In many cases, either the employer or the employee in this situation would be expected to supervise the AI as a human (or organization) in the loop and would have difficulty successfully suing any of the upstream entities.

In addition to determining who is responsible, these various configurations also demonstrate that multiple bodies of law are involved in the event of harm occurring, such as traditional negligence, products liability, or contracts (including warranty considerations, consumer obligations, and limitations on liability). Contracts are typically executed between each entity

---

privity with respect to tort liability. Benjamin C. Zipursky, *Rights, Wrongs, and Recourse in the Law of Torts*, 51 VAND. L. REV. 1, 32-34 (1998).

[115] The use of AI for purposes of inclusion in a company's technology stack has most recently been dubbed "AI as a Service." Jennifer Cobbe & Jatinder Singh, *Artificial Intelligence as A Service: Legal Responsibilities, Liabilities, and Policy Challenges*, 42 COMP. L & SEC. REV., 2021, at 3-7.

and the entities next to it, creating a complex web of rules affecting who is liable and when. As described in Part I, technology stacks used to create AI systems also involve a variety of third-party technology providers, all of which can potentially affect the safety and efficacy of AI systems. These third-party technology providers are also subject to contractual provisions.[116]

The prevalence of AI startups means that AI will likely be licensed by a broad number and variety of organizations in many sectors or be acquired by larger organizations. For acquired AI, although organizations may have purchased confidential information and intellectual property about the AI, organizational staff may not have the requisite knowledge and skillsets to understand the AI or its limitations. For example, AI could be used in a healthcare application it was not trained for, diagnosing prostate cancer, despite being trained to diagnose breast cancer.

Organizational licensees will likely have even less information available to them if the licensing organization wishes not to share trade secrets or confidential information, including information affecting safety, efficacy, or fairness, particularly when a general AI application is being used in different contexts.[117] For example, navigational AI used for

---

[116] In many other contexts, such as cybersecurity and privacy, I use the terminology "sub-privity" to illustrate that organizations in contractual relationships with a user have countless contractual relationships behind it that affect its ability to full obligations in the contractual relationship with a downstream user. Although a downstream user would not have privity in contract with any of these subcontractors or licensors, they may have a relationship as construed under tort, however remote. While the economic loss doctrine could foreclose tort actions in the immediate contract, some jurisdictions may not extend that to subcontractors outside the immediate contract. *See supra* Section II.A; *see, e.g.*, Chapman Custom Homes, Inc. v. Dallas Plumbing Co., 445 S.W.3d 716 (Tex. 2014) (finding that issues of privity between a homeowner and a subcontractor of a housing contract did not also bar recovery under the economic loss doctrine due to the existence of a contract between the homeowner and contractor).

[117] Start-ups may maintain a great deal of confidential information and intellectual property as the basis for attracting investors and to discourage competitors. The high expense of securing a patent combined with a questionable ability to enforce the patent against infringers has likely reduced interest in patenting (and associated public disclosure). Many start-

automated driving could also be used for farm machinery, drone operation, pharmaceutical delivery robots in a hospital, or food delivery. The potential risks for each of these contexts could differ dramatically: getting the wrong food delivered is categorically different from taking the wrong medication or using a combine on the wrong wheat field.[118] The data used to train AI in each of these contexts would also be very different and could result in greater or less risk depending on how the AI is trained. The nature of these relationships and relative access to information is important when determining who should be liable and under what circumstances.

B. Existing Tort-Law Conventions Create AI Liability Challenges

The tort system is ill-prepared for the raft of AI litigation to come. AI will likely frustrate key goals of tort law: (1) to provide compensation for the injured party (the value of what was lost or taken), (2) to remedy the injury while punishing a tortfeasor, and (3) to incentivize the defendant (and other would-be tortfeasors and future wrongdoers) not to commit a tort in the future.[119] In economic approaches to tort law, the

---

ups are relying on confidentiality and trade secrecy as an alternative. Disclosing these details, then, could destroy this status and competitive advantage.

[118] Andrew Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315, 1344 (2021) (quoting Ryan Calo and explaining the dynamics of context-shifting AI and challenges from the perspective of a data scientist). As Selbst notes, an individual using AI may also be unable to determine what reasonable operation or use of AI could be, in the moment, in part because the user may not understand whether an AI tool is designed to replicate human operation or exceed it. Andrew Selbst calls this the "category of foreseeability," illustrating that what could be foreseeable risk in one category may be completely unforeseeable in another. Indeed, "the [AI] designers . . . deny they had any idea this would happen . . . they did not in their wildest nightmares imagine it would kill people." *Id.* at 1344 n.34 (quoting Ryan Calo).

[119] *See* Mark Geistfeld, *Negligence, Compensation, and the Coherence of Tort Law*, 91 GEO. L.J. 585, 585 (2003). As Geistfeld acknowledges, rationales under the Restatement (Third) of Torts, which dominantly focus on remedying an injustice, are not consistent with dominant theories that

goal of tort is also to distribute costs most efficiently when such wrongs occur.[120]

Tort law is based on relational interactions, actual or constructive, that leave one person or entity harmed.[121] The relational construct of tort law demands the law figure out who is to blame and to what extent they are to be blamed.[122] Tortfeasors may commit torts, but plaintiffs may not always be compensated as a result.[123] Ultimately, tort law is not designed to protect all people from all injuries committed by a tortfeasor; the ability of a plaintiff to prevail is tightly tied to their ability to satisfy each tort's specific prima facie case requirements and to survive a defendant's affirmative defenses. Under negligence theories, plaintiffs are also expected to exercise reasonable caution to protect themselves. In a variety of tort theories, including contributory negligence, comparative negligence, failure to warn, or user error, an AI user could be foreclosed from a remedy.[124]

---

apply to other forms of tort, such as strict liability, which can incorporate a more deterrence-based approach focusing on wrongful conduct, rather than focusing on the injury as justification. *Id.* at 586.

[120] Guido Calabresi & Jon T. Hirschoff, *Toward a Test for Strict Liability in Torts*, 81 YALE L.J. 1055, 1060 (1972). Although several modern theoretical models have been explored in tort, Calabresi's nomenclature of "cheapest cost avoider" has stood the test of time.

[121] *See* John C.P. Goldberg & Benjamin C. Zipursky, *Torts as Wrongs*, 88 TEX. L. REV. 917, 946 (2010).

[122] *Id.*

[123] *Id.* at 944.

[124] Contributory and comparative negligence involve the weighing of "wrongfulness" with respect to various parties involved in the tort, including the plaintiff. *See* Cristina Carmody Tilley, *Tort Law Inside Out*, 126 YALE L.J. 1320, 1332-34 (2017). Additionally, torts may include situations where plaintiffs consent to a tort, fail to exercise reasonable care to protect themselves, or assume the risk contractually. *See id.* at 1344; *see also* Aaron D. Twerski & James A. Henderson Jr., *Fixing Failure to Warn*, 90 IND. L.J. 237, 250-52 (2015) (describing a plaintiff's obligation to argue an alternative warning in the event the plaintiff alleges the warning was insufficient as a basis for a failure-to-warn claim when plaintiffs do not follow or notice the warning); Robert A. Adler & Andrew F. Popper, *The Misuse of Product Misuse: Victim Blaming at Its Worst*, 10 WM. & MARY BUS. L. REV. 337, 346 (2019); David G. Owen, *Products Liability: User Misconduct Defenses*, 52 S.C. L. REV. 1, 2 (2000).

When situations involving AI result in harm to people or other entities,[125] plaintiffs will use some variety of tort claims to establish liability, most often negligence or products liability (manufacturing defect, design defect, or communication), depending on who is providing or using the technology.[126] Other tort claims could apply depending on the AI's principal function and context for its use, such as malpractice or libel.[127] This Section discusses the most likely torts applicable to safety and efficacy cases and how plaintiffs may have difficulty recovering from developers outside the loop.

1. Embodied and Virtual Products

One of the first considerations for plaintiffs is who to sue and under what theory that entity or individual can be sued. The characterizations of a defendant and their relationship to the plaintiff complicates this inquiry. For AI, one of the biggest challenges is whether the AI is "embodied," a term that can

---

[125] One consideration for torts involving AI in general will be whether or not a plaintiff can adequately plead reasonably cognizable, sufficiently specific, non-speculative and past or recently impending injury. Although many AI harms may result, tort law may only be successfully used for harms that can be considered injuries the law will compensate. Harms involving unauthorized data use, data loss, or other privacy harms may require some reconceptualization of harm as injury. This Article does not aim to solve this much broader debate. For more information, see, for example, Daniel J. Solove & Danielle Keats Citron, *Risk and Anxiety: A Theory of Data-Breach Harms*, 96 TEX. L. REV. 737, 747-53 (2018) (describing a need for expansion of our conceptions of harm regarding data use); Daron Acemoglu, *Harms of AI* 1-3 (Nat'l Bureau Econ. Research, Working Paper No. 29247, 2021) (describing the types of harms AI could create, some of which could be compensable injuries and some that would not).

[126] *See* Victor E. Schwartz, *The "Restatement (Third) of Torts: Products Liability": A Guide to Its Highlights*, 34 TORT & INS. L.J. 85, 88-90 (1998).

[127] See John D. Banja, Rolf Dieter Hollstein & Michael A. Bruno, When Artificial Intelligence Models Surpass Physician Performance: Medical Malpractice Liability in an Era of Advanced Artificial Intelligence, 19 J. AM. COLL. RADIOLOGY 816, 816 (2022); Scott Schweikart, Who Will Be Liable for Medical Malpractice in the Future? How the Use of Artificial Intelligence in Medicine Will Shape Medical Tort Law, 22 MINN. J.L. SCI. & TECH. 1, 3 (2020); Eugene Volokh, Large Libel Models? Liability for AI Output, 31 J. FREE SPEECH L. 489, 492 (2023) (describing the potential for libel lawsuits and acknowledging a current OpenAI lawsuit).

mean implanted in a physical body or within a physical space.[128] Embodied AI is likely to be considered a "product," while other AI, even AI performing the same functions, may be characterized as a service.

Because software (including AI) may be embodied or not,[129] an "AI device" could be physical or virtual.[130] Because many AI systems direct physical functions or provide inputs into some physical processes, they may legally function like a physical product without actually being embodied. Virtual systems can cause harm to a variety of people and organizations when they direct physical function, even if they are not actually within a physical device.[131] The ultimate characterization of AI can have an enormous effect on which tort theories can be pursued. Virtual systems or services will most likely be subject to negligence claims, whereas embodied systems or products will most likely be subject to products liability claims. I discuss each in turn.

2. Negligence

For AI that is purchased as software or an application, negligence is likely to be the cause of action against people or organizations who later harm a person or entity using that AI. A negligence prima facie case requires that a plaintiff successfully argue that the defendant breached their reasonable duty of care, and that the breach is both the factual and proximate (legal) cause of the plaintiff's injury. Most cases will look something like this:

---

[128] Tehseen Zia, *Embodied AI: How It Bridges the Gap Between Mind and Matter*, TECHOPEDIA (Sept. 15, 2023), https://www.techopedia.com/embodied-ai-bridging-the-gap-between-mind-and-matter [https://perma.cc/XM87-Y98E] (describing the "embodiment" of AI in physical devices).

[129] *See* Bryan H. Choi, *Crashworthy Code*, 94 WASH. L. REV. 39, 65-66 (2019) (describing software's intangibility, leaving it to debate whether it is a good or a service depending on its embodiment).

[130] Charlotte Tschider, *Regulating the IoT: Discrimination, Privacy, and Cybersecurity in the Artificial Intelligence Age*, 96 DENV. L. REV. 87, 97 (2018) (describing the use of AI in conjunction with cloud services and Internet-connected devices).

[131] These systems can direct physical function and operate virtually. *See* Choi, *supra* note 129, at 79.

> *Casperillo Foods is a food manufacturing company using AI to manage food safety in its plants. The AI connects to robotics machinery that tests product temperature and pH at various points in the cycle. Casperillo has developed its system in-house but purchased a ready-made AI software product to direct the machine's functions, since none of the technology developers on staff have AI experience. Although human operators oversee the manufacturing process and periodically hand-test the products, the AI tool and its human operators failed to catch a few products that were undercooked and made customers sick. Now, Casperillo Foods has issued wide recalls of its products at the direction of the Food and Drug Administration (FDA) and is facing a class-action negligence suit for its food safety AI.*

In this example, Casperillo Foods did not create the AI—it only purchased and used it. Here, the human operators supervising manufacturing are "in the loop." This configuration is a common AI use case: an organization or person uses AI directly or indirectly, but the AI is not created by them. The outcome of this interaction can create potential torts. Part of what negligence law has to determine is who should be held responsible for the injuries of this class, and to what extent each party might be liable. Negligence requires the plaintiff to demonstrate that the defendant breached their duty of care and that the breach was the proximate (legal) cause of their injury-in-fact. Should Casperillo, who uses the AI but did not create it, be more liable than the AI developers who created it?

In order to demonstrate that a duty of care has been breached, plaintiffs must first establish what the duty of care for a reasonably prudent organization using AI or a reasonably prudent AI company should have been. Then, they must demonstrate that the defendant did not fulfill that reasonable

duty.[132] Duty of care and breach are contextual relative to the specific defendant, the case, and the established injury.[133] For example, a defendant using AI might be expected to perform a different duty of care than the company licensing an AI software product. The concepts of duty of care and breach are tied to foreseeability—performing one's duty of reasonable care supposes that any party in a similar situation would have considered foreseeable risks and acted reasonably with respect to them.[134] The challenge with AI, of course, is that courts have not yet established what reasonable duties actually look like—that is, whether and to what extent some parties may be required to do more to protect individuals and entities that use AI.

As Andrew Selbst rightly notes, negligence law is designed to create legal responsibility "for and only for such harm as [the defendant] could reasonably have foreseen and prevented."[135] Foreseeability, then, provides limits on what an organization is required to do and when an organization has breached some duty. Foreseeability is especially significant for fact patterns that involve the intervening actions of a third party that interrupt the causal chain leading to the harm at issue. Where some action could have prevented the intervening action from occurring, such as closing a cybersecurity vulnerability, it may be tempting to find, without thorough examination, that the defendant breached their reasonable duty.[136] However, when a

---

[132] Rebecca Crootof, *The Internet of Torts: Expanding Civil Liability Standards to Address Corporate Remote Interference*, 69 DUKE L.J. 584, 611 (2018).

[133] *Id.* at 627.

[134] *Id.* at 615.

[135] *See* Selbst, *supra* note 118, at 1332 (quoting H.L.A. Hart and Tony Honoré). Negligence-based torts require a showing of foreseeability, but strict liability doctrines do not. "[t]he plaintiff is no longer required to impugn the maker, but he is required to impugn the product." John F. Scarzafava, *An Analysis of Products Liability Defenses in the Aftermath of Hopkins*, 9 ST. MARY L.J. 261, 263 (1977) (quoting Dean Page Keeton); Chris Lewis, *The Need for a Legal Framework to Regulate the Use of Artificial Intelligence*, 47 U. DAYTON L. REV. 285, 304 (2022).

[136] Complex technologies could create potential issues for determining a breach of reasonable duty. For AI, courts may determine that an organization may be in breach simply because the technology failed—not that the protective measure was reasonable, and the defendant failed to

third party's intervening actions, or even a technology's failures are not foreseeable, finding a breach of duty may be unreasonable.[137] This contemporary limitation on negligence serves to narrow the field of potential defendants. Coupled with tort reform efforts, limits on foreseeability aim to improve efficiency and prevent excesses.

AI creates unusual challenges for negligence law's foreseeability requirement. Because the range of what *could* happen in an AI system is more expansive than in non-AI systems, the possibility of harm resulting from something unexpected is higher.[138] Despite this, AI in some cases are closed systems,[139] meaning that they are not designed to be open to an external party's change or manipulation.[140] While to some degree AI learn and change from collective inputs of users, the underlying decisional system and its attendant affordances and constraints should not.[141] In most

---

implement it. *See* Tschider, *supra* note 93, at 77-78 (describing attacks where even objectively reasonable security practices can be compromised).

[137] Negligence and products liability actions limit recovery to injuries that are reasonably foreseeable, including how such injuries were caused and who was injured. The tort system limits the scope of potential recovery. *See* Stephen D. Sugarman, *Why No Duty?*, 61 DEPAUL L. REV. 669, 690-93 (2012).

[138] *See* Selbst, *supra* note 118, at 1362; Ryan M. Calo, *Open Robotics*, 70 MD. L. REV. 571, 596-97 (2011).

[139] Will Douglas Heaven, *The Open-Source AI Boom Is Built on Bigtech's Handouts. How Long Will It Last?*, MIT TECH. REV. (May 12, 2023), https://www.technologyreview.com/2023/05/12/1072992/the-download-open-vs-closed-ai-and-googles-uneasy-demo [https://perma.cc/8RPH-J6QG].

[140] This does not mean that all AI systems are fully protected from outside manipulation, however. AI systems may still be subject to cyberattacks and similar unauthorized access, despite being designed as closed systems. *See generally* Marcus Comiter, *Attacking Artificial Intelligence*, HARV. UNIV. BELFER CTR. PAPER (Aug. 2019), https://www.belfercenter.org/sites/default/files/2019-08/AttackingAI/AttackingAI.pdf [https://perma.cc/X9FV-SWAK] (describing the many types of attacks targeting all AI systems).

[141] Affordances and constraints create limits to users: affordances show a user what they can do, while constraints limit what a user can do. These are

manufacturing contexts, including for coding and development companies, the manufacturer is "the expert as to the product," which means very few potential issues are unforeseeable, "unless [they are] also unknowable."[142]

Although there are legitimate challenges of foreseeability in AI, the reality is that a lawsuit will be dismissed when the cause of the injury-in-fact derives from an unforeseeable causative chain. For example, if an AI system suddenly functions in a completely unforeseeable way, generally that injury will not be recoverable. Foreseeability also matters a great deal when intervening causes affect the outcome. Cyberattacks are an existing risk for AI systems, but it is not always clear which attackers could attack and what attack vectors they might use. In these cases, it might be difficult for a plaintiff to argue that the cyberattack used was foreseeable to the extent that the AI creator should have anticipated and prevented it.[143]

One way to overcome the foreseeability issue is to adopt a positional approximation of foreseeability rather than a narrow conception of it. It is becoming more known that AI can

---

crucially important for AI that have human users. Sam Stone, *The Design of Everyday (AI) Things,* MEDIUM (Oct. 6, 2023), https://towardsdatascience.com/the-design-of-everyday-ai-things-26516d928566 [https://perma.cc/FH94-972S]. *See also* Deirdre K. Mulligan & Helen Nissenbaum, *The Concept of Handoff as a Model for Ethical Analysis and Design*, *in* THE OXFORD HANDBOOK OF ETHICAL AI 232, 239-40 (2020); Dirk John Pons & John Kenneth Raine, *Design Mechanisms and Constraints*, 16 RSCH. ENG'G DESIGN 73, 83 (2005) (calling for shift from solution creation to constraint generation).

[142] Data scientists and engineers create AI systems that, even to them, are inscrutable. Moreover, as a means of protecting investments, disclosable information may instead be protected as a trade secret or other proprietary information and maintained as confidential. This means that even the creators closest to AI products have limited knowledge as to its function, and what information is available will generally not be shared downstream. *See* Tschider, *supra* note 54, at 699, 706 (quoting Daniel Susser). *See also* William E. Westerbeke, *The Sources of Controversy in the New Restatement of Products Liability: Strict Liability Versus Products Liability*, 8 KAN. J.L. & PUB. POL'Y 1, 6 (1998).

[143] Anat Lior, *Artificial Intelligence and Tort Law—Who Should be Held Liable when AI Causes Damages*, HEINRICH BÖLL STIFTUNG (Dec. 24, 2021), https://il.boell.org/en/2021/12/24/artificial-intelligence-ai-tort-law-and-network-theory-who-should-be-held-liable-when-ai [https://perma.cc/9EUU-7JP4].

function in unexpected (and often dangerous) ways, so foreseeability may not be as useful in these cases.[144] One way to incorporate foreseeability involves a positional approximation of access to information (i.e., who has the most information) to identify who is most likely to discover potential issues before they occur. Given that downstream organizations are likely to have less information (and therefore limited foreseeability regarding potential harms), it would make most sense to attribute harm to the AI creator that formalized AI design choices and sold the AI, knowing licensees would use it for certain purposes.[145] Given that an AI development company is likely to know more about how the AI functions and its potential risks because they created it, they are better positioned to accept liability than organization licensees—or humans in the loop. Although this does not mean that organizations creating AI or their employees will *always* foresee certain outcomes, it does mean they are comparatively better positioned than product customers or users to anticipate harm—and avoid it.

---

[144] For example, Baris Soyer and Andrew Tettenborn examine this problem from the European perspective, finding that under the Consumer Protection Act 1987, defective products are framed such that "the safety of the product is not such as persons generally are entitled to expect," rather than foreseeability, leaning towards strict liability. Baris Soyer & Andrew Tettenborn, *Artificial Intelligence and Civil Liability—Do We Need A New Regime?*, 30 INT'L J.L. & INFO. TECH. 385, 392 (2022). Andrew Selbst recognizes that foreseeability as currently used would all but foreclose most plaintiff's cases unless AI is made interpretable or explainable. *See* Selbst, *supra* note 118, at 1362.

[145] Foreseeability is a difficult hurdle but may be an extension, as Benjamin Zipursky offers, of the wrong itself. *See* Benjamin C. Zipursky, *Foreseeability in Breach, Duty, and Proximate Cause*, 44 WAKE FOREST L. REV. 1247, 1274 (2009). If we center on the breach of the duty of care established in statute, for example, and the AI development company that ultimately makes the choices that shape the features and function of the AI then knowingly sells such an AI to industries to use it in particular ways, it would be difficult to argue that potential harms are unforeseeable, at least in relation to other potential defendants.

In negligence cases involving computer systems like AI, causation also presents potential issues.[146] Although plaintiffs have the distinct challenge of demonstrating foreseeability, they will likely face similar challenges proving causation due to the unintelligibility of AI decisions: information about what actually happened and *why* may not be understandable or persuasive to judges and juries, and approximations of decisions may not truly explain what happened.[147] For example,

---

[146] Powers & Ganascia, *supra* note 58, at 46. Describing Nancy Cartwright's account of randomized controlled trials in computer-based statistical analysis as illustrating how statistical analysis in AI can only "vouch for" a scientific claim, rather than "clinching" it. Overall, as Judea Pearl asserts, "one cannot substantiate causal claims from associations alone." In both cases, causation is challenged both because of the unintelligibility of inscrutable systems but also because much of AI functions through intelligent generalization and proxy.

[147] Explanation is often positioned as an important function of AI systems, and it has been positioned as key tool for regulating AI. *See, e.g.*, Danielle Keats Citron, *Technological Due Process*, 85 WASH. U. L. REV. 1249, 1253-55, 1296-97, 1308-13 (2007) (proposing explanation of the extent a decision relies on automated decision-making, the actual code used for decision-making, public review and involvement in system design, and rigorous testing); Frank Pasquale, *Restoring Transparency to Automated Authority*, 9 J. TELECOMM. & HIGH TECH. L. 235, 236-40, 244-50 (2011) (describing the specific circumstances where transparency may outweigh secrecy); Price, *supra* note 52, at 429, 432-37, 440-42, 460, 465-66 (introducing the unique impacts of algorithmic decision-making in the healthcare sector and the need to validate predictive analytics); PASQUALE, *supra* note 52, at 160-62 (describing a model of 'qualified transparency' applicable to search engines); Adam Thierer, Adrea O'Sullivan & Raymond Russell, *Artificial Intelligence and Public Policy*, MERCATUS RSCH. 36 (Aug. 23, 2017), https://www.mercatus.org/publications/artificial-intelligence-public-policy [https://perma.cc/APX9-LMCE] (describing the availability of tools for transparency and the potential for regulatory involvement). As a result, alternative explanatory approximations have been explored, such as using an auditing or supervisory AI to analyze the AI's decisions or developing summarized forms of AI decisions using natural language processing. *See* François Candelon, Theodoros Evgeniou & David Martens, *AI Can Be Both Accurate and Transparent*, HARV. BUS. REV. (May 12, 2023), https://hbr.org/2023/05/ai-can-be-both-accurate-and-transparent [https://perma.cc/9A8T-EFYT]. The role of explanation could be important after an AI has made a decision that results in significant potential harm, such as factory robotics physically injuring a human worker, or AI interview software turning down all candidates with a greater concentration of melanin in their skin. The ability to actually explain a specific decision

if an AI system stops functioning, and someone is injured, plaintiffs may have difficulty proving what the cause of the malfunction was. Although for some aspects of the AI system, such as server performance, databases, or other infrastructure, the defendant may possibly determine why a harm occurred, plaintiffs will still face difficulty discovering it for themselves.

If AI models are inscrutable, information collected about them, even reasonable interpretations, may not provide definitive information about what actually happened, at least not easily or accurately. For these reasons, in pretrial motions that permit the court to evaluate the plaintiff's claims in a light most favorable to them, courts could adopt a relaxed approach to the prima facie case showing. For example, in motions to dismiss and motions for summary judgment, courts should allow plaintiffs to illustrate how AI likely caused their injury-in-fact based on the AI's functionality (from the perspective of the plaintiff) rather than actual knowledge of the AI's failure. Since AI will largely be inscrutable and other information about the AI may be confidential, relaxing the causation requirement would make sense until after the trial court's discovery phrase.

3. Products Liability

When negligence actions do not apply, products liability actions likely will, though many of the same impediments for plaintiffs bringing negligence claims may apply here as well. Because AI can be used in disembodied and embodied products, the dichotomy between negligence and products liability will need to be resolved by courts. As described in this Section, products liability addresses the harms of AI products better than traditional negligence, though it requires a

---

leading to some harm, however, may be difficult to provide and may not be terribly useful. I have previously suggested alternative models for interrogating the black box, which under most circumstances do not require explanation. *See* Tschider, *supra* note 54, at 689, 695-96, 707, 719 (questioning the utility of explanations and proposing a registration system for AI, including hosting the live algorithm for testing).

reconceptualization of how products liability applies to AI in the first place.

A typical products-liability case involves the plaintiff seeking compensation from the manufacturer or seller for injuries caused by a product's design or manufacturing defect, or inaccurate or misleading communication about the product.[148] Manufacturers are not responsible for all potential product-related injuries, only those arising from defects and failures to warn about potential risks.[149] Defects render products unsafe for their intended use, encompassing product flaws, packaging, labeling, and advertising.[150] If a product functions as intended and is safe for regular use, it is not considered defective. This model creates some issues when applied to AI because AI is designed to change and, in some cases, is designed to be licensed and used for multiple purposes.[151]

One of the most challenging aspects of the products-liability model is its narrow application to some, but potentially not all, AI products.[152] As discussed *supra* in Section II.B.1, AI products may be virtual rather than "tangible personal property distributed commercially for use or consumption," and "services, even when provided commercially, are not products."[153] This clear delineation, shared by the Uniform Commercial Code's § 2-105, is very narrow in its application, and would apply only to AI models that can meet the definition of "product."[154] Despite this, many AI products are a combination of intangible and embodied aspects. Intangible products often affect physical function but direct that function virtually. Often, companies sell intangible AI as a product.

---

[148] *See* Crootof, *supra* note 132, at 654.
[149] *Id.*
[150] *Id.*
[151] *See* Ryan E. Long, *Artificial Intelligence Liability: The Rules Are Changing*, STAN. CTS BLOG (Mar. 17, 2023), https://cyberlaw.stanford.edu/blog/2023/03/artificial-intelligence-liability-rules-are-changing-1 [https://perma.cc/4MGX-FXWS].
[152] In many cases, computer software is considered a service rather than a product. *See* JACOB TURNER, ROBOT RULES: REGULATING ARTIFICIAL INTELLIGENCE 95-96 (2019).
[153] RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 19 (AM. L. INST. 1998).
[154] U.C.C. § 2-105.

Practically, this means that AI of various kinds would be subject to significantly different liability models. Embodied AI would likely fall to products liability in design negligence and manufacturing defects (which impose strict liability), along with the Uniform Commercial Code's Section 2 (governing goods).[155] Intangible AI, however, would be subject to general negligence and the Restatement of Contracts, both of which establish different liability conventions.

In the abstract, this distinction offers clarity about which liability path plaintiffs might take, but in reality, many AI products cross these boundaries depending on their use. AI products that may be intangible when created may be used to direct embodied products later. In other cases, things typically not considered products, like real property, may be redrawn as products when the facts support it.[156] The most recent Restatement of Torts acknowledges the flexibility of a products liability classification, noting that products liability applies to:

> Other items, such as real property and electricity . . . when the context of their distribution and use is sufficiently analogous to the distribution and use of tangible personal property that it is appropriate to apply the rules as stated in this Restatement."[157]

---

[155] William A. Tanenbaum, *Theories of AI Liability: It's Still About the Human Element*, REUTERS (Sept. 20, 2022), https://www.reuters.com/legal/litigation/theories-ai-liability-its-still-about-human-element-2022-09-20 [https://perma.cc/965F-D93C]; Michael D. Scott, *Tort Liability for Vendors of Insecure Software: Has the Time Finally Come?*, 67 MD. L. REV. 425, 434-36 (2008).

[156] For example, courts have reconsidered these lines in the past when specific technologies cross into the products category from another area, such as nuclear power plants (typically considered real property) that are legally considered products. *AI for Nuclear Energy*, AI FOR GOOD (2023), https://aiforgood.itu.int/event/ai-for-nuclear-energy [https://perma.cc/YFG3-N9UW].

[157] RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 19 (AM. L. INST. 1998).

Therefore, products liability can be extended to situations that involve non-traditional products,[158] but most courts may not extend it to intangible products, except when products are a blend of tangible and intangible components.[159] Importantly, the origin of the distinction between products and services under the Second Restatement, adopted in 1965, was that services were historically between two parties, whereas products were mass-produced.[160] Today, that historic distinction does not reflect reality. Yet, in all cases, software and computer system products continue to be construed under the Restatement[161] (and the UCC) as services, rather than products.[162]

In the event that courts choose to extend products liability to software, the Restatement appears to anticipate and be

---

[158] The question of what is considered a "product" has been debated for a long period of time, including prior to the passage of the Third Restatement. Charles E. Cantu, *The Illusive Meaning of the Term "Product" Under Section 402A of the Restatement (Second) of Torts*, 44 OKLA. L. REV. 635, 638 (describing court analyses of product using a policy-based approach).

[159] *See* Bexis, *New Decision Directly Addresses the "Is Software a Product" Question*, DRUG & DEVICE L. (May 2, 2022), https://www.druganddevicelawblog.com/2022/05/new-decision-directly-addresses-the-is-software-a-product-question.html [https://perma.cc/NM42-ZQRG] (acknowledging the primacy of tangibility in products liability rather than negligence but describing the court's conclusion in *Corley v. Stryker Corp.* that "not only the physical implant but also the defendant's 'software used in creating it' was a 'necessary part' of the 'product' as a whole").

[160] *See* Cantu, *supra* note 158, at 640. Indeed, many efforts nearly all efforts to extend strict liability to services have failed. *Id.*

[161] The Restatement (Third) of Torts recognizes three discrete bases for products liability suits: design defect and failure to warn (which are usually not considered strict liability), and manufacturing defect (which is usually considered strict liability). This move reduced, to some extent, the position that products liability is "strict liability" law—as framed predominantly under the Restatement (Second) of Torts, which incorporated a law and economics perspective. In contrast, the Restatement (Third) of Torts introduced a more compensation-oriented model. *See* Geistfeld, *supra* note 119, at 585.

[162] Where claims involve a mix of services and products, courts adopt a predominant purpose test, similar to that used to distinguish between goods and services in contracts cases. *Id.* at 643. Where "component parts" become part of the finished product, typically these will be construed as a product for purposes of liability. *Id.* at 656.

reasonably amendable to a broader interpretation.[163] Software categories that are considered "functional" are more likely to be considered products, rather than custom-made or purely expressive software.[164] Courts in a variety of circuits, including the Ninth Circuit,[165] have identified software as products, but the cumulative result is still highly speculative. However, in many Circuits, there is no bright-line rule today. There are a variety of reasons why at least *some* AI could be considered a product: it may function as a product (directing the functionality of something physical or embodied), and it may be marketed like a product. Adopting a products-liability scheme could enable courts to examine traditional aspects of the AI product, in particular the degree to which these AI products can withstand and mitigate harm when they fail.[166] Even without an embodiment, virtual AI products could still cause physical harm indirectly. Consider medical AI: AI is routinely used today for diagnostic purposes, and it is used to streamline a wide variety of medical processes.[167] Here, too, there are inconsistencies in what counts as a product. Although the FDA classifies diagnostic software designed for use in a

---

[163] David Berke, *Products Liability in the Sharing Economy*, 33 YALE J. REGUL. 603, 611 (2016); *c.f.* Choi, *supra* note 129, at 53 n.62.

[164] Berke, *supra* note 163, at 613-16.

[165] At least one recent case has demonstrated the Ninth Circuit's willingness to apply products liability to virtual products, such as social media sites. *See In re* Social Media Adolescent Addiction/Personal Injury Products Liability Litigation, No. 22-md-03047, 2023 WL 7524912, at *19, 20 (N.D. Cal. Nov. 14, 2023); *see also* Holbrook v. Prodomax Automation Ltd., No. 17-cv-219, 2021 WL 4260622 (W.D. Mich. Sept. 20, 2021); Quinteros v. InnoGames, No. C19-1402, 2022 WL 898560 (W.D. Wash. Mar. 28, 2022), *recons. denied*, No. C19-1402RSM, 2022 WL 953507 (W.D. Wash. Mar. 30, 2022), and *aff'd in part, rev'd in part and remanded*, No. 22-35333, 2024 WL 132241 (9th Cir. Jan. 8, 2024).

[166] *See* Choi, *supra* note 129, at 110-11 (describing fault tolerance in software, including courts examining code).

[167] *See* Terrence Mills, *AI for Health and Hope: How Machine Learning Is Being Used in Hospitals*, FORBES (Feb. 16, 2022), https://www.forbes.com/sites/forbestechcouncil/2022/02/16/ai-for-health-and-hope-how-machine-learning-is-being-used-in-hospitals [https://perma.cc/T5QE-CG8Q].

medical setting a "medical device" that is manufactured,[168] such software would likely not meet the definition of a "product" for purposes of products liability. Whether AI will fall under a negligence or products-liability scheme is currently an unsettled topic. However, as explained through the rest of this Part, there are advantages to embracing a cohesive products-liability model for AI.

4. Design and Manufacturing Defects

Even if products liability could be consistently used for all AI regardless of its embodiment, there are other challenges in applying the doctrine. Chief amongst these is the legal separation of design and manufacture, which are different processes in traditional product manufacturing. In traditional product manufacturing, a design document is handed off for product manufacturing once prototypes have been rendered and tested.[169] However, computer programming, especially in its modern form, combines design *and* development in largely iterative processes.[170]

---

[168] *See* Sara Gerke, 'Nutrition Facts Labels' for Artificial Intelligence/Machine Learning-Based Medical Devices—The Urgent Need for Labeling Standards, 98 GEO. WASH. L. REV. 79, 92, 122 (2023) (describing 'Software as a Medical Device'). In fact, the FDA published four conditions that must be met to not be considered a medical device. Id. at 103. Although in some cases, medical device manufacturers would be protected from common law claims in products liability or negligence, in many cases, device manufacturers would potentially be subject to products liability or negligence claims. See Tschider, supra note 111, at 1580 n.145.

[169] *See Design for Manufacturing: Importance and Key Principles*, RAPID DIRECT (Nov. 11, 2021), https://www.rapiddirect.com/blog/why-design-for-manufacturing-is-important-in-prototyping [https://perma.cc/TMC7-4KS4].

[170] Clark Savage Turner, Software and Strict Products Liability: Technical Challenges to Legal Notions of Responsibility 38, 41 (1999) (Ph.D. dissertation, U.C. Irvine) (http://users.csc.calpoly.edu/~csturner/master.pdf) [https://perma.cc/A2LB-QCLE]. In the "Code and fix" and the "Waterfall with feedback" software development models, a developer codes but makes changes based on conformance to a design document, an inherently iterative process. *Id.* Computer programming has only continued to become more iterative and cyclic, where the final result is very close to portions of the design process, so much so that tweaks and changes in design can occur up to the very last steps. The result has become an amalgamation of design and development practices. *See* Irving Wladawsky-Berger, *Using Agile*

As described in Part I, AI's development process, at least in its initial stages, consists of continuous learning, testing, and tuning.[171] Although the basics of AI could be "designed," it is not designed in the same way a conventional product is designed, which is fixed.[172] For locked systems, the AI only becomes fixed after it has completed all of its testing and tuning and is stamped for release. Therefore, design and manufacture converge into one, at least for functional software or applications that power tangible products and direct how they function.[173] But despite the convergence of the two categories in the AI context, "design defect" and "manufacturing defect" are accompanied by different legal models. Design defects generally are held to a negligence standard requiring proximate causation, while a manufacturing defect is governed by strict liability.[174]

Depending on how certain activities are described, one court could determine that a defendant is strictly liable, while another could determine another plaintiff cannot sue the defendant under a manufacturing defect standard. Unlocked, dynamic AI systems pose a greater challenge because they are designed to learn continuously as they are being used, which means that design and manufacture (and the potential defects introduced) never end.[175] Rather, the only way to mark some

---

Processes to Develop AI-Based Solutions, WALL ST. J. (Aug. 23, 2019), https://www.wsj.com/articles/using-agile-processes-to-develop-ai-based-solutions-01566568928 [https://perma.cc/X2Y3-A33Q].

[171] *See supra* Part I.

[172] *See* RAPID DIRECT, *supra* note 169.

[173] *See* Turner, *supra* note 170, at 41.

[174] RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 (Am. L. Inst. 1998).

[175] Dynamic inscrutability means that even if the decisions of a model could be pinpointed, their continuous evolution and change would make it difficult to demarcate what is "design" and what is "production," which means the AI is working in a live environment. *See generally* Mahardhika Pratama, Andri Ashfahani & Edwin Lughofer, *Unsupervised Continual Learning via Self-Adaptive Deep Clustering Approach*, ARXIV 48 (June 28, 2021), https://arxiv.org/abs/2106.14563 [https://perma.cc/H7EU-6G8C] (describing the challenge of developing accurate continuous learning algorithms, especially for deep learning models and avoiding 'catastrophic forgetting').

completion is the AI's availability for sale or use on the open market, though this completion would only serve as a legal fiction for the purposes of retrofitting product-liability models onto AI products.

Manufacturing defects historically meant defects introduced in the manufacturing process that could harm an unsuspecting person downstream. In contemporary software, it is expected that issues will arise and that organizations will continuously monitor and resolve these issues within a reasonable time period. This has been largely managed through regular updates to close security vulnerabilities, fix software bugs, or offer new features, distributed through wireless connectivity.[176]

Although a straightforward defect-management model could be used, AI licensed for use and personalization to organizations as part of their technology stack could create different problems. For example, AI could be created and trained in an initial form by an AI startup, then licensed.[177] While that version of the AI model may be free from defects, it could then be used as a basis for organizational AI that has a particular use.[178] As discussed, AI could be used for any number of purposes.[179] If harm results from that use, it may be difficult to pinpoint whether the original AI introduced defects or whether the application of the AI to a specific context created them.[180]

---

[176] John Villasenor, *Products Liability Law as a Way to Address AI Harms*, BROOKINGS INST. (Oct. 31, 2019), https://www.brookings.edu/articles/products-liability-law-as-a-way-to-address-ai-harms [https://perma.cc/52ZT-JAHQ].

[177] *See supra,* Section II.A.

[178] An additional challenge is that usually AI software and hardware creators avoid liability unless the products were defective when made. Long, *supra* note 151.

[179] Sara Miteva, *Choosing Between Open Source and Closed Source Coding*, MEDIUM (July 12, 2019), https://medium.com/wearelaika/differences-between-open-source-and-closed-source-code-7d3427b3acc0 [https://perma.cc/GNR9-TSQG].

[180] As Ryan Abbott notes, "By creating a framework for loss shifting from injured victims to tortfeasors, tort law deters unsafe conduct." In the event that tort law is frustrated by the nature of AI technologies, it fails to realize its goal of deterrence. *See* Ryan Abbott, *The Reasonable Computer: Disrupting the Paradigm of Tort Liability*, 86 G.W. L. REV. 1, 3 (2018).

Overall, products liability is recognized at the state level, and despite the American Law Institute (ALI) completing the Restatement (Third) of Torts: Products Liability in 1997, states still have a wide variety of frameworks they use in this context, including or excluding failures to warn, breaches of implied warranty, or strict liability.[181] The inconsistency in how products liability is being applied could mean that AI development companies, their licensees or acquirers, and AI users will not know when they are acting reasonably or not, exposing them to potential legal risk.

In AI, defective design could lead to downstream safety issues. In products liability, however, it isn't enough to claim that the design of the AI is defective. Rather, a plaintiff must show that "the foreseeable risks of harm posed by the product could have been reduced or avoided by the adoption of a reasonable alternative design."[182] This requirement presumes that a plaintiff has adequate information to pose some reasonable alternative design. The practical challenge for a plaintiff, however, is having enough information to actually propose these alternatives (or to provide an expert witness), which often adds time and expense to a lawsuit.[183] Adding further complexity is a requirement in some jurisdictions to propose a reasonable alternative design that is sufficiently distinct and different from the challenged design, both in design and in function.[184]

The reasonable alternative design also incorporates a foreseeability risk-utility balancing test: "whether a reasonable alternative design would, at a reasonable cost, have reduced

---

[181] Brenda Leong & Jey Kumarasamy, *Third-Party Liability and Product Liability for AI Systems*, IAPP (July 26, 2023), https://iapp.org/news/a/third-party-liability-and-product-liability-for-ai-systems [https://perma.cc/6H5N-KHCQ].

[182] *See* Schwartz, *supra* note 126, at 88.

[183] Katherine A. Kelter, *Limits on "Reasonable Alternative Design" in Product Defect Cases*, AMER. BAR ASS'N (Apr. 24, 2017), https://www.americanbar.org/groups/litigation/committees/products-liability/practice/2017/limits-on-reasonable-alternative-design-product-defect-cases [https://perma.cc/Y3ZL-E5RF].

[184] *Id.*

the foreseeable risk of harm posed by the product and, if so, whether the omission of the alternative design by the seller . . . rendered the product not reasonably safe."[185] The Restatement (Third) of Torts: Products Liability at least in theory created some flexibility in establishing liability for manufacturing defects. One principal contribution was the creation of a form of strict liability for manufacturing defects: a product "contains a manufacturing defect when the product departs from its intended design even though all possible care was exercised in the preparation and marketing of the product."[186] Under the Restatement, plaintiffs arguing strict liability do not have to demonstrate independent reasonable duties, including whether or not the plaintiff's injury was foreseeable. Of course, the Restatement (Third) of Torts is not the law in all states, though it aims to create consistency in legal standards across the United States.

Manufacturing defects, on the other hand, occur *after* a design has been completed while the product is being manufactured, and these defects are the primary basis for strict liability.[187] AI, however, may pose a challenge here because its "design" could also be construed as "manufacturing." Limiting manufacturing-defect claims to only physical housing and functions of a device would likely not encapsulate a wide variety of AI defects manifesting a physical function. Although theoretically, it might seem like AI defects that wouldn't fall under manufacturing would instead fall under design defect, many of these defects might not become known until the AI is manufactured and used as part of the device, especially if the AI is unlocked and continuously learning in the real world.

For both manufacturing defects and design defects, plaintiffs injured by AI could benefit from a relaxation of evidentiary standards. One treatise, describing the ALI's position in the Restatement (Third) of Torts, acknowledges the imprecision of determining products liability:

---

[185] Spencer H. Silvergate, *The Restatement (Third) of Torts Products Liability: The Tension between Product Design and Product Warnings*, 75 FLA. BAR J. 10, 10 (2011) (quoting Comment d to § 2).

[186] RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 (AM. L. INST. 1998).

[187] As Rebecca Crootof notes, in products liability cases, strict liability is typically reserved for manufacturing defects rather than labeling or design decisions causing injury. *See* Crootof, *supra* note 132, at 623 n.211.

> [P]roof of a product defect *may be made circumstantially, without proof of a specific defect*, when the incident causing the harm was of a kind that would ordinarily occur only as a result of a product defect and was not, in the particular case, solely the result of causes other than a product defect existing at the time of sale or distribution.[188]

If plaintiffs are able to use circumstantial information to support their claims, plaintiffs may survive motions to dismiss and summary judgment to proceed to discovery with a manufacturer plaintiff. Because plaintiffs may have substantially less information available to them, especially for AI products, relaxed evidentiary standards in pretrial motions could increase the potential viability of their claims.

Common defense strategies for escaping strict liability for manufacturing defects include arguing that the product was modified after it left the manufacturer's custody or that the user misused the product.[189] In some jurisdictions, a user may have assumed risk either implicitly or explicitly, or the user may share in the risk in a comparative negligence jurisdiction.[190] If labeling and other communication do not adequately warn of potential hazards in using AI or otherwise mislead users about the AI system's operation, communication may also be the basis for a products-liability lawsuit.[191]

Courts use various tests to determine whether there is liability for design defects, such as the consumer expectation test, where product danger may be actionable if it's beyond ordinary consumer expectations.[192] This test presumes some

---

[188] AM. L. PRODS. LIAB. § 1:4, Westlaw (3d ed. database updated May 2024) (emphasis added).

[189] These limitations could potentially justify the need for products-liability causes of action that address newer forms of products-liability issues, such as torts involving remote interference for Internet-connected products. *See* Crootof, *supra* note 132, at 623.

[190] *Id.* at 634.

[191] 38 A.L.R.5th 683 (1996).

[192] 73 A.L.R.5th 75 (1999).

degree of warning was not present when danger was beyond consumer expectations, but it is only used in a minority of states.[193] The majority approach is the risk-utility test, which considers whether the risk was high enough to require investment in preventing safety hazards.[194]

5. Failure to Warn and Assumption of Risk

In addition to design and manufacturing defects, the third type of products-liability claim is failure to warn. In failure to warn cases, organizations have not effectively communicated potential risks so that buyers could make an informed decision.[195] In situations where a person knowingly exposes themselves to the risk of using AI despite being warned of potential risks, assumption of risk could apply. In its most basic form, assumption of risk is accepting any risk that results from using or doing something inherently unsafe. The theoretical basis for assumption of risk is that an individual, after receiving adequate risk information, voluntarily accepts potential risks (either impliedly or expressly), which could include consent or constructive awareness of potential risks through observation or warning.[196] For this reason, assumption of risk is classically used in situations where risk cannot be meaningfully reduced for an individual, and the individual proceeds anyway.[197] Whether that assumption of risk is valid depends on whether a person would have reasonably expected that kind of harm. For

---

[193] *Id.*; Sunghyo Kim, *Crashed Software: Assessing Product Liability for Software Defects in Automated Vehicles*, 16 DUKE L. REV. 300, 305-06 (2018).

[194] David G. Owen, *Design Defects*, 73 MO. L. REV. 291, 299 (2008); *see* Kim, *supra* note 193, at 307-08.

[195] It should be noted that products liability did not always encompass a strict-liability approach. Historically, English courts embraced *caveat emptor*, or "buyer beware," to reflect a free-market approach. *See* Abbott, *supra* note 180, at 14.

[196] Although courts have shifted to a reasonableness analysis rather than a heavy focus on consensual risk taking, consent is still relevant to assumption of risk analysis (as well as duty/breach). *Id.* Indeed, "those offended by the deeds of others cannot complain about that which, to an important extent, is the upshot of their voluntary undertaking." Avihay Dorfman, *Assumption of Risk, After All,* 15 THEORETICAL INQUIRIES L. 293, 294 (2014). Both private contract and tort involve allocation of risk. *Id.*

[197] *See* Dorfman, *supra* note 196, at 294. Indeed, a powerful defense to intentional torts is consent. *Id.*

example, a child breaking an arm on a football field might be expected, but the same child being hurt by a downed electrical wire might not be.

Some courts have folded assumption of risk into duty/breach analysis, liability apportionment (resulting in either partial defendant liability or no liability in comparative and contributory negligence jurisdictions, respectively),[198] and contractual liability waivers.[199] Indeed, the shift from contributory negligence to comparative negligence, where calculation of the percentage of liability is common, has challenged courts to determine whether assumption of risk doctrine will even survive.[200] Still, others have maintained assumption of risk analysis as a substantive category.[201]

Although a defendant may use express or implied assumption of risk as an affirmative defense to negligence claims (arguing that they legally disclosed risks, so they cannot be held responsible for harms disclosed), this is not a universally effective defense.[202] Specifically, many courts engaged in assumption of risk analysis find that a person cannot agree to risks that are broadly construed.[203] Typically, a defendant needs to demonstrate that the plaintiff knew of the dangerous condition, knew it was actually dangerous,

---

[198] Apportionment can be related to who bears the burden of proof, especially for plaintiffs that have no ability to prove that their injury can be attributed to one or more tortfeasors. *See* Choi, *supra* note 129, at 95.

[199] Kenneth W. Simons, *Reflections on Assumption of Risk*, 50 UCLA L. REV. 481, 483 (2019).

[200] *Id.* at 486.

[201] *Id.* at 483.

[202] It should be noted that two theoretical models apply to this analysis: the first, the "redundancy objection," subsumes assumption of risk into duty analysis; the second, the "analytical objection," argues that an assumption of risk should never be a complete bar to liability. Under both models, assumption of risk is not absolute. Dorfman, *supra* note 196, at 295. In the second case, consent to risk exposure is not sufficient; rather, consent should be with respect to material risk. *Id.* A third conception of risk takes an economic, markets-based approach. *Id.* at 296 n.6.

[203] *See* Simons, *supra* note 199, at 483.

appreciated the extent of the danger, and voluntarily exposed themselves to it.[204]

Whether it would be reasonable to enforce an assumption of risk turns on reasonable behavior, including whether warnings are present, and whether the plaintiff acted reasonably with respect to them. Similar to reasonable duty, assumption of risk "operates on the correlativity of the right to safety held by a potential victim and the duty of care owed to her by a potential injurer."[205] For example, an AI navigation app user that becomes injured after following a dangerous walking path certainly has some responsibility to protect themselves, say, by turning around if a path intended to slope towards a steep embankment. This behavior would be consistent with a reasonable person's behavior. If the person continued down the path, they may have assumed a risk of injury. However, the user probably does not carry all of the responsibility for whether or not they took a path marked legitimate in the first place, and most courts seem to agree.[206] For example, if there is an alternate, safer path available, most people will take it. Sometimes this means that assumption of risk hinges on whether a safer path could have been available, be it a trail or an AI feature.[207]

Liability shifting through an assumption of risk affirmative defense must be justified, meaning that if liability shifts, it must be fair to do so.[208] In many cases where AI is used, there are multiple issues that challenge an assumption of risk defense to negligence. First, AI use (and assumption of risk associated with it) may not actually be voluntary because a company using AI is not obligated to disclose it is doing so.[209] Second, the observability of potential risks and knowledge of material harms is likely to make any assumption of risk, even with express consent, unfair. Because an AI user may not be able to fully understand the potential harms even if they are stated on

---

[204] Amy L. Stein, *Assuming the Risks of Artificial Intelligence*, 102 B.U. L. REV. 979, 987 (2022).
[205] *See* Dorfman, *supra* note 196, at 307.
[206] *See* Simons, *supra* note 199, at 508 n.77.
[207] *See id.* at 509-10.
[208] *See* Dorfman, *supra* note 196, at 315.
[209] *See* Stein, *supra* note 204, at 1008-10 (explaining that different settings could exhibit different levels of voluntariness).

a label or within a warranty document (because of AI's overall opacity), AI users could not be assuming the actual risk.[210] The user would have to expend a great deal of time and money to evaluate the reliability and safety of an AI application, and users occupy a poor position to make that evaluation, relative to the organization that created or licensed the AI. Although the AI might be inscrutable to both parties, the defendant is in a relatively less costly position to avoid or mitigate potential AI risk.

Overall, products liability has the potential to work effectively in AI contexts, but it must evolve to anticipate new inconsistencies, evidentiary challenges, and ways in which would-be defendants will capitalize on assumption of risk through extensive and excessive warnings. Without a new model for evaluating these potential claims and defenses, it is likely that inconsistency and inefficiency in the courts will result.

C. Humans in the Loop May Be (Human) Liability Shields

Despite the likely role of negligence and products liability to AI safety claims, organizations will likely rely upon limitations established through private ordering to avoid liability altogether. Historically, manufacturers have used contract law to reduce the likelihood of plaintiff recovery in negligence and products-liability lawsuits. All parties involved in AI transactions routinely enforce contracts between them, including formal written contracts, contracts of adhesion that include warranty details in a consumer product's box,[211] or terms of use agreements. These contracts fulfill multiple

---

[210] *See id.* at 1005.

[211] For all goods, warranties of merchantability and fitness for a particular purpose are implied (unless overtly disclaimed) in every contract. U.C.C. § 2-314; U.C.C. § 2-315. These implied warranties can be useful in the event that a product does not behave as expected, causing damage for which an individual has to pay. Organizations, however, can disclaim these, so long as they follow U.C.C. § 2-316 and state requirements for how to disclaim them. *Id.* The United States has no federal law that pertains to consumer rights with respect to warranties outside of the U.C.C.'s provisions as incorporated in state law.

purposes: they establish contractual terms, limit or transfer liability, establish expectations for use, and serve information-sharing functions.

For some AI, contracts establish additional requirements a human must fulfill in conjunction with the AI's function, or else the person may lose warranty coverage and potential recovery in tort.[212] In short, a combination of contractual terms may establish expectations of a human in the loop, then use the human's failures as a basis for transferring liability from the AI-development company or its downstream successors to the human in the loop. Although this Article does not discuss the details of warranty exclusions and limitations,[213] AI development companies often use contracts and contractually referenced documents, such as manuals and labels, in conjunction with limited liability provisions, to limit their liability.[214] Functionally, these provisions often limit risk, so long as they are not written too broadly and the contract is not deemed unconscionable.

An excellent example of defendants using these strategies is the automobile manufacturer Tesla, which is known, at least in part, for its Autodrive feature. Tesla has been the subject of at least four lawsuits involving contractual escapes from liability involving a "human in the loop," but these lawsuits are not likely to succeed despite the serious injuries of Tesla drivers and bystanders.[215] In these cases, injured or killed Tesla owners

---

[212] Warranty limitations are used heavily in software. *See* Scott, *supra* note 155, at 437-38.

[213] Warranty exclusions are one area where AI-development companies are beginning to substantially limit liability based on breach of warranty or breach of implied warranties. *See* Villasenor, *supra* note 176. Although these warranty exclusions can be connected to broader concerns of liability, they often present as breach of contract actions, rather than those in tort. Ka-on Li & Jeffrey Jones, *Be Wary of Warranties for Software Design*, JONES DAY (Aug. 2018), https://www.jonesday.com/en/insights/2018/08/be-wary-of-warranties-for-software-design [https://perma.cc/E7MF-AZYN]. For that reason, this Article does not address the details of those potential claims, though details of a warranty may come in as evidence of unreasonable behavior on the part of a plaintiff, just as other documentation could also be introduced.

[214] *See* Leong & Kumarasamy, *supra* note 181.

[215] One plaintiff, Justine Hsu, sued Tesla for strict product liability, negligence, breach of duty to warn, breach of implied warranty, intentional

(or their estates) alleged that defects in Tesla's autopilot software feature (updated remotely by Tesla) were the proximate cause of injuries when the feature:

- Swerved into a curb, causing the airbag to deploy, injuring the driver by fracturing the driver's jaw, knocking out teeth, and causing facial nerve damage
- Drove under the trailer of a semi-truck that had pulled into the road, shearing off the roof and killing the driver
- Caused a driver to veer off a highway at 65 miles per hour and strike a tree, causing the car to burst into flames, killing the driver and severely injuring an 8-year old passenger
- Exited a freeway, ran a red light, and crashed into another car, killing its two occupants[216]

In these cases, Tesla described the driver's actions as the proximate cause of the resulting accidents. Under Tesla's Autodrive manual, drivers have the obligation to maintain control of the vehicle and monitor its activity. Tesla's user manual and website overtly direct drivers, among other directives, to "ensure all cameras are clean," that "it is the

misrepresentation, and common law fraud concealment. Complaint at 1, Hsu v. Tesla, Inc., 2023 WL 8168177 (Cal. Super. May 14, 2020) (No. 20ST-cv-18473). None of Hsu's claims were successful. *See* Abhirup Roy, Dan Levine & Hyonjoo Jin, *Tesla Wins Bellwether Trial Over Autopilot Car Crash*, REUTERS (Apr. 22, 2023, 5:43 AM EDT), https://www.reuters.com/legal/us-jury-set-decide-test-case-tesla-autopilot-crash-2023-04-21 [https://perma.cc/A464-9S76].

[216] *See, e.g.*, Dani Anguiano, *Landmark Trial Involving Tesla Autopilot Weighs if 'Man or Machine' at Fault*, THE GUARDIAN (Nov. 14, 2022), https://www.theguardian.com/technology/2022/nov/14/tesla-autopilot-landmark-case-man-v-machine [https://perma.cc/8DSB-WAE3]; Jonathan Stempel, *Tesla is Sued by Drivers Over Alleged False Autopilot, Full Self-Driving Claims*, REUTERS (Sept. 14, 2022), https://www.reuters.com/business/autos-transportation/tesla-is-sued-by-drivers-over-alleged-false-autopilot-full-self-driving-claims-2022-09-14 [https://perma.cc/42WL-57CR]; Trisha Thadani, Faiz Siddiqui, Rachel Lerman & Jeremy B. Merrill, *Tesla Drivers Run Autopilot Where It's Not Intended—with Deadly Consequences*, WASH. POST (Dec. 10, 2023, 6:00 AM EST), https://www.washingtonpost.com/technology/2023/12/10/tesla-autopilot-crash [https://perma.cc/94A5-WXHZ].

driver's responsibility to stay alert, drive safely, and be in control of the vehicle at all times," that a driver must "always watch the road in front of you and be prepared to take corrective action at all times," that "Autosteer is a hands-on feature" and a driver must "keep [their] hands on the steering wheel at all times . . . and always be prepared to take immediate action," and to "never depend on Autosteer to determine an appropriate driving path."[217]

Even the Full Self-Driving setting (at the time of writing, in beta testing), is accompanied by seemingly contradictory statements to the feature's designed use of "full self-driving":

> Always remember that Full Self-Driving (Beta) . . . does not make Model 3 autonomous and requires an attentive driver who is ready to take immediate action at all times . . . Failure to follow these instructions could cause damage, serious injury or death. It is your responsibility to familiarize yourself with the limitations of Full Self-Driving (Beta) and the situations in which it may not work as expected . . . there are numerous situations in which driver intervention may be needed.[218]

These obligations, while enclosed in a user manual and not drafted in a formal contract, are referenced in warranty documentation and limited liability language, which are part of the contract. Collectively, Tesla has prevented most lawsuits from being successful principally based on requiring a human to correct the AI-enabled car's failures.[219]

---

[217] *Model 3 Owner's Manual*, TESLA (2023), https://www.tesla.com/ownersmanual/model3/en_us/GUID-E5FF5E84-6AAC-43E6-B7ED-EC1E9AEB17B7.html [https://perma.cc/9W6A-4Z7Y].

[218] *Id.*

[219] *See* Andrew J. Hawkins, *Tesla Wins Another Court Case by Arguing Fatal Autopilot Crash Was Caused by Human Error,* THE VERGE (Oct. 31, 2023), https://www.theverge.com/2023/10/31/23940693/tesla-jury-autopilot-win-liable-micah-lee [https://perma.cc/UK39-NC2X]; Jay Peters, *Tesla Wins Lawsuit that Blamed Autopilot for Crash,* THE VERGE (Apr. 21, 2023), https://www.theverge.com/2023/4/21/23693482/tesla-lawsuit-blamed-autopilot-crash [https://perma.cc/L6PY-KGJ9]. In an interesting turn of

As in the Tesla cases, many limitations on liability are enforced, including limitations on consequential and punitive damages. These limitations on liability may be subject to invalidation if (1) the law outright bars them (e.g., the defendant is committing fraud or the limitations are contrary to public policy), or (2) they are unconscionable.[220] However, it is rare for these provisions to actually be invalidated. Limitations on liability have generally escaped challenge based on unconscionability or violating public policy, at least in cases where drafting was not overly broad or poorly executed.[221]

Limitations on liability apply to a variety of would-be plaintiffs. AI companies licensing or selling AI products to organizations likely will include similar limitations to those in consumer contracts, or they may include indemnification provisions in the event the purchasing party is using the AI product as part of their technology stack.[222] It is a smart strategy

---

events, Tesla customers have brought new lawsuits alleging Tesla misled purchasers that the Autodrive packages were safe. *See, e.g.*, Fred Lambert, *Tesla is Now Facing a Class Action Lawsuit Over 'Misleading' Claims About Self-Driving*, ELECTREK (Sept. 15, 2022, 8:13 AM PT), https://electrek.co/2022/09/15/tesla-facing-class-action-lawsuit-over-misleading-claims-self-driving [https://perma.cc/XK5G-9AGT].

[220] Emlin McClain, *Contractual Limitation of Liability for Negligence*, 28 HARV. L. REV. 550, 554-55 (1915); Joey McCue & Logan Wells, *Limitation of Liability Clauses and Unconscionability*, 25 S.C. BAR J. 16, 18-19 (2013).

[221] *See, e.g.*, Negrete v. Citibank, N.A., 187 F. Supp. 3d 454, 469 (S.D.N.Y. 2016) (describing limitations of liability provisions as part of the negotiated allocation of the risk of economic loss and enforceable); Core-Mark Midcontinent, Inc. v. Sonitrol Corp., 300 P.3d 963, 968 (Colo. Ct. App. 2012) (finding limitations of liability generally enforceable due to its allocation of risks in a bargained-for exchange); Rosenberg v. Cape Coral Plumbing, Inc., 920 So. 2d 61, 64 (Fla. Dist. Ct. App. 2005) (finding limitations of liability enforceable so long as they are clear); LDCircuit, LLC v. Spring Cmmc'ns Co., L.P., 364 F. Supp. 2d 1246, 1268 (D. Kan. 2005) (finding limitations on liability enforceable unless they are not illegal, unconscionable, or contrary to public policy).

[222] Bill Molloy, John Shirley & Philip Tansley, *Tech Disputes: Exclusions and Limitations of Liability*, LEXOLOGY (Sept. 7, 2021), https://www.lexology.com/library/detail.aspx?g=a15988fc-6415-4ca3-a67f-b99d875b99c2 [https://perma.cc/5HMM-WM75] (arguing that limitations on liability are interpreted less strictly than exclusions and liability caps, but

on the part of AI manufacturers to insulate themselves, but these strategies all but foreclose recovery for individuals who may not be in a meaningful position to avoid these injuries.[223]

When defendants cannot be held liable, plaintiffs ultimately absorb the cost of harm. The idea that "nobody" is liable is inaccurate: in a tort system, if the defendant is not required to make the plaintiff whole, the plaintiff absorbs the relative costs of their injury. Madeleine Clare Elish famously named these scenarios, where machine or computer error is incorrectly attributed to humans, as "moral crumple zones," due to their highly damaging impact with very little ability to hold agents accountable.[224] As Elish notes, complex cases involving multiple errors can "obscure the responsibility of key individuals" involved in developing the technology.[225]

Legal obscurity creates different legal challenges, depending on the jurisdiction. First, plaintiffs will face difficulty proving negligence without any statutory duty of care and without enough information to demonstrate causation. Next, plaintiffs may be foreclosed altogether from bringing a products liability claim based on the fact their AI (for example, disembodied AI) is not considered a "product." In the event they can demonstrate its status, design defects could be difficult to prove because the AI system itself is inscrutable, may not have a reasonable alternative design, and the AI itself could just be part of a tech stack, licensed from another organization. A manufacturing defect, typically enjoying strict liability, would likely be challenging to argue, since AI malfunctions are not likely caused by errors in manufacturing assembly (when the physical component is 'manufactured').

All of these claims may not even be possible to examine in the event an AI manufacturer has created significant

---

that all may be used in technology contracts, especially between sophisticated entities).

[223] Limitations on liability are not completely unrestricted, but they are likely to be enforced unless the "underlying transaction 'affects the public interest.'" Food Safety Net Servs. v. Eco Safe Sys. USA, Inc., 147 Cal .Rptr. 3d 634, 642 (Ct. App. 2012) (describing limited liability provisions as enforceable unless the underlying transaction affects the public interest).

[224] Madeleine Clare Elish, *Moral Crumple Zones Cautionary Tales in Human-Robot Interaction*, 5 ENGAGING SCI., TECH. & SOC'Y 40, 41 (2019).

[225] *Id.* at 51.

obligations for humans to complete, where their failure triggers a limitation of liability provision that forecloses these claims. In short, it is highly likely that the least knowledgeable people and organizations, and in many cases, humans in the loop, will ultimately absorb the liability of AI creators and even licensees, shielding them from liability.

D. AI Challenges Historical Liability Models

While cases involving AI failures *could* be examined using existing liability models, including contractual limitations on liability, these models will not create an efficient, consistent system, and in many cases the results will be unfair to plaintiffs. By doubling down on a model that will not work effectively for AI, both organizations and plaintiffs risk losing trust and certainty in the legal process: organizations may not know how to avoid liability, and plaintiffs may presume any harm suffered will not be compensated. Legal uncertainty may, over time, affect consumer purchasing and investment in innovation. In addition to these legal uncertainties, certain factors, including information asymmetry and automation bias, affect the foundations upon which existing liability models are built. These factors dramatically affect the overall fairness of shifting liability to a human in the loop.

As previously described, the various choices AI creators make for design, data, and infrastructure dramatically affect the safety, efficacy, and fairness of AI systems. However, the orientation of various downstream businesses and consumers reduces the degree to which information is effectively dispersed, impacting if these parties can be expected to reasonably avoid risk or understand when AI is not performing as designed. Essentially, humans in the loop, often employees or consumers charged with supervising their work or purchasing products, will have the least amount of information available about the AI itself. Although information asymmetry will be a general problem for most downstream entities, humans in the loop are perhaps the least well-positioned in terms of bargaining power to learn anything about how the AI actually functions.

Humans without skill or access to proprietary information cannot readily evaluate other aspects of system choices due to information asymmetry.[226] To some degree, AI information asymmetry is intractable—even with all of the information scientists have about the AI, users might not overcome AI inscrutability and intelligibility issues to truly challenge its function. Information asymmetry usually occurs in relational constructs between organizations and consumers, governments and residents, organizations and employees, or in fiduciary relationships. In relationships involving information asymmetry, typically one party (which has more skill or information) has some special duty or increased obligation to the other party, or the secondary party is relieved of some duty.[227]

Information asymmetry plays out in a variety of ways in contracts, often requiring disclosures or providing copies of information to the individual with less information. For example, a seller in real estate transactions are often required to disclose "material defects related to the condition of the property."[228] These mandated disclosures promote fair business transactions through equal access to otherwise hidden and unavailable information (or information that would require substantial time investment from a buyer to uncover).[229] Even more, the timing of such disclosures could be crucial to their effectiveness.[230] AI, for its part, is not subject to any special disclosures of any kind, though often warnings are

---

[226] Michael Guihot, Anne F. Matthew & Nicolas P. Suzor, *Nudging Robots: Innovative Solutions to Regulate Artificial Intelligence*, 20 VAND. J. ENT. & TECH. L. 385, 421-22 (2017).

[227] For example, fiduciary relationships often involve information asymmetry. Claudia Haupt, *Platforms as Trustees: Information Fiduciaries and the Value of Analogy,* 134 HARV. L. REV. F. 34, 36 (2020). *See also* Neil Richards & Woodrow Hartzog, *Taking Trust Seriously in Privacy Law*, 19 STAN. TECH. L. REV. 431, 451, 457 (2019) (describing the importance of trust in all relationships involving data collection and use).

[228] Amanda C. Brown, *Caveat Emptor, Buyer Boo-ware*, 77 TEX. BAR J. 780, 781 (2014).

[229] *See, e.g.*, Maria Maciá, *Mandatory Disclosure for Ethical Supply Chain: Market Responses to Conflict Mineral Reports*, 13 HARV. L. REV. 189, 190 (2023) (describing the use of mandatory disclosure regimes as an alternative to direct regulation).

[230] Stephanie Stern, Temporal Dynamics of Disclosure: The Example of Residential Real Estate Conveyancing, 57 UTAH L. REV. 57, 58 (2005).

used in contracts, warranties, and user manuals to notify users of high-level risks and their obligations in avoiding them. However, these documents do not usually include meaningful information about the AI and how it actually functions to help users, including humans in the loop, avoid potential harms.

As described extensively in Part I, the design, training, testing, and data selection choices made by AI developers have a direct impact on the safety and efficacy of AI systems.[231] Despite this, objectively good AI technology choices do not necessarily presume that a human can readily determine how the AI is making decisions or usefully interrogate them.[232] In fact, in some cases, safer AI could result in less of an ability to interrogate, interrupt, or require it to explain itself.[233] Some human goals require AI that is complex, opaque, and inscrutable to accomplish them. Forcing disclosure of AI details could create greater risk of compromise, hypervigilance,

---

[231] *See supra* Section I.A.

[232] The amount of information available to a human user may be fairly limited based on an organization's desire to protect its inventions. Because most inventions may not be eligible for protection under patent law, the ability to make money from them lies instead in non-disclosure, such as confidentiality of proprietary information and potentially trade secrecy. *See* David L. Schwartz & Max Rogers, *"Inventorless" Inventions? The Constitutional Conundrum of AI-Produced Inventions*, 35 HARV. J.L. & TECH. 531, 564 (2022). This means that even if a human in the loop sought more information about the AI, it might not be available to them, or only approximations of the information might be available.

[233] *See* Sabine N. van der Veer et al., *Trading off Accuracy and Explainability in AI Decision-Making: Findings from 2 Citizens' Juries*, 28 J. AM. MED. INFORMATICS ASS'N 2128, 2129 (2021). It is unknown how certain users could react to differences in accuracy versus explainability, but at least from initial indications, accuracy may be more important than explainability in some contexts, while in others, explainability may be essential. This seems to suggest that understanding the relevant needs downstream would inform design to facilitate explainability where it is comparatively more necessary. *See id.* at 2136 (describing the findings of a UK study comparing healthcare and non-healthcare contexts and finding that citizens' juries ranked accuracy over explainability in healthcare contexts, but not in others).

or gaming, which could compromise AI safety and efficacy.[234] Ultimately, disclosure of AI details (assuming they could be made available) to improve human-in-the-loop knowledge may not actually improve the performance of humans in the loop or lead to fair liability outcomes.

Although providing extensive upfront information may not be tremendously useful, courts could address information asymmetry in the litigation process. If the defendant, whether AI creator or licensee integrating AI into a product, is in a relatively better position to discover information that could affect AI safety, efficacy, or fairness than a plaintiff, the law could impose more burden on the defendant than the plaintiff to compensate for a legally cognizable injury. Arguably, a defendant development company who creates or acquires the AI is in a comparatively superior position to gather information, such as design, training, testing, and tuning information about the AI, than a downstream business or consumer user. Even if reasonable warning communications and labels are made available to these users (e.g., as described in the Tesla cases), a difference in the skillset needed to understand the information and take appropriate steps to avoid risk could significantly burden these individuals.[235] For example, courts could determine what information should require production in discovery requests regardless of confidentiality or trade secrecy claims (with appropriate accompanying judicial protection) based on (1) AI system or product details, (2) the degree to which the organization was involved in the AI's design choices, and (3) the degree of control over the AI in its live form. For organizations that have little to no knowledge or control, discovery requests could be similarly limited. This approach might also help plaintiffs determine who should be potential defendants. Figure 2 illustrates the many layers (and associated distance) between

---

[234] Andrew Burt, *The AI Transparency Paradox*, HARV. BUS. REV. (Dec. 13, 2019), https://hbr.org/2019/12/the-ai-transparency-paradox [https://perma.cc/DE5E-9J47].

[235] A lack of warning labels or effective warning labels can often be the basis for design-defect liability. *See* Kim, *supra* note 193, at 308-09. However, it is hard to conceptualize a label for AI that can effectively guard against potential AI risks, which are typically the product of complex functions and may be difficult to foresee, even from the perspective of an AI developer.

original AI creators and harmed plaintiffs and defendants most likely to provide relevant information.

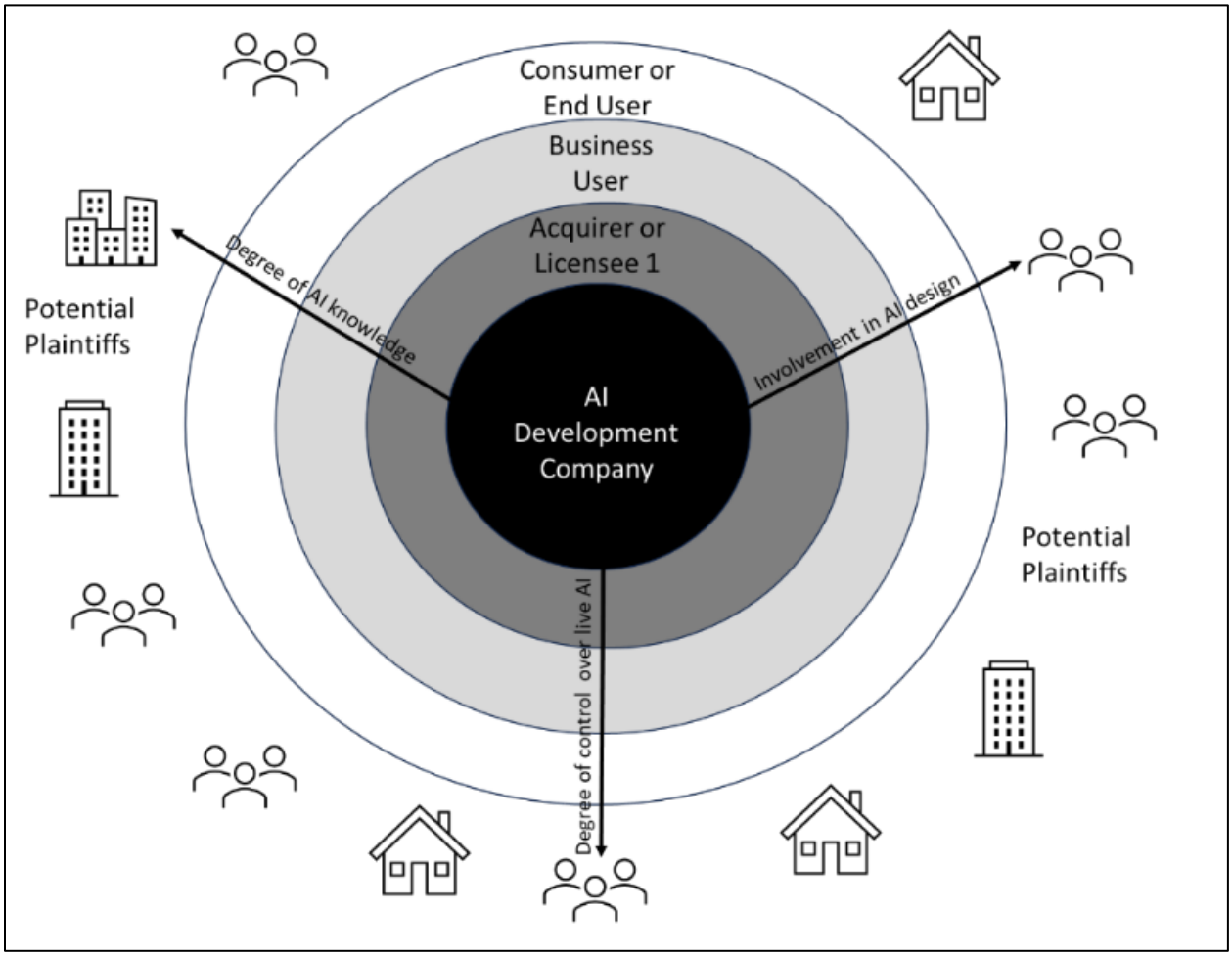

Figure 2: Degree of distance between AI developers and potential plaintiffs

In addition to information asymmetries and issues of information dispersion, automation bias prevents humans in the loop from effectively challenging technology. Automation bias is a type of cognitive bias: systematic, universally occurring

inclinations that distort thinking.[236] When humans are biased, their decisions are inaccurate, suboptimal, or wrong, even if they result in generally acceptable outcomes.[237] Human biases are, to some degree, unavoidable; our brains simply exhibit a mismatch between evolutionary heuristics and an environment's current context.[238]

Humans do not always perceive task complexity objectively. Complexity is usually viewed as an extension of how difficult or time-consuming a task is for a human to complete. However, humans and computers simply work differently, "for us it is much more difficult to multiply two random numbers of six digits than to recognize a friend in a photograph."[239] When a human observes a task completed by a computer that the human perceives as complex, the human (1) is not able to perform the task because it is complex, and (2) believes the computer must have done it correctly because of its complexity. Tasks that are easy for the brain are not necessarily simple for a computer to perform, and what may be complex for the brain can be simple for a computer to perform; this is called the "Moravec's Paradox".[240]

Humans beat out computers in a variety of ways, however. Humans are excellent at perceiving whole objects faster than portions of whole objects.[241] Words are perceived better than letters.[242] Our motor skills and perception are nearly automatic.[243] These are difficult characteristics for computers. Humans are also better at handling unexpected situations and making decisions about them as a flexible team.[244] Despite automation bias, these differences may explain why we perceive humans in the loop as effective computer supervisors. Overall, humans and computers think and function, respectively, differently.

---

[236] *See* J.E. (Hans) Korteling et al., *Human-Versus Artificial Intelligence*, 4 FRONTIERS A.I., Mar. 25, 2021, at 1, 4.
[237] *Id.*
[238] *Id.* at 5.
[239] *Id.* at 6.
[240] *Id.*
[241] *Id.*
[242] *Id*.
[243] *Id.*
[244] *Id.* at 7.

Automation bias may cause humans to accept a computer-generated outcome without critically examining it.[245] Typically, automation bias results from situations where a human does not have the ability to scrutinize how a system generated a certain outcome.[246] Although automation bias has existed as long as computers have been around, AI seems to exacerbate it because AI incorporates a more complex decisional process.[247] While some humans may defer to AI, others may actually assume it will perform poorly in comparison to a human, even when it provides objectively better results; this reaction is known as "algorithmic aversion".[248]

Whether bias or aversion, neither mental model accurately reflects AI function or promotes effective human in the loop behavior. Moreover, these humans in the loop will not be able to reasonably absorb shifted liability because they cannot fulfill the task they are expected to perform. Liability shifting may occur either when evaluating proximate cause or through contract, and courts have the opportunity to determine the degree to which a human in the loop should absorb liability for malfunctioning AI. Where liability shifting is established through contract, courts should deem these clauses unenforceable under unconscionability or public policy objections, depending on their breadth and context.[249]

Contractual liability shifting usually occurs when someone is using the AI for personal purposes or for their business. For

---

[245] *See* Stefan Strauß, *Deep Automation Bias: How to Tackle a Wicked Problem of AI?*, BIG DATA & COGNITIVE COMPUTING, Apr. 2021, at 6-8.
[246] *Id.*
[247] *Id.*
[248] S. Mo Jones-Jang & Yong Jin Park, *How Do People React to AI Failure? Automation Bias, Algorithmic Aversion, and Perceived Controllability*, J. COMPUTER-MEDIATED COMMC'N, Jan. 2023, at 1, 2.
[249] Unconscionability is a combination of both substantive and procedural unconscionability, and therefore is a case-by-case determination. RESTATEMENT (SECOND) OF CONTRACTS § 208 (AM. L. INST. 1981). Certain limitations on liability can leave consumers with little ability to seek compensation for serious injuries and may be deemed unconscionable or against public policy. *See* Joshua N. Cohen, *Sound the Alarm: Limitations on Liability in Alarm Service Contracts*, 85 FORDHAM L. REV. 813, 826-27 (2016).

example, a manufacturing company licenses AI used for robotics applications to make their factory function more efficiently. A liability shifting provision might state something like:

> *In the event of any unanticipated function causing injury or harm, ABC Manufacturing shall hold XY Robotics harmless against any subsequent direct or third-party claims resulting from non-performance of activities specified in the XY Robotics manual, warranty materials, or this Agreement.*[250]

If the manual described in the limitation on liability provision states the human has to supervise and intervene actively in the application's direction of robotics equipment, it is fairly likely that any harm that results will be attributed to the human's failure to perform as specified. For courts illustrating why these provisions should be deemed unconscionable or unenforceable on the grounds of public policy, at least three justifications could bar liability shifting through limitations on liability and other provisions.

First, opacity issues and automation bias combine to significantly disadvantage a human in the loop. Because a human brain is not designed to understand machine-based complexity quickly (or at all), a human cannot effectively supervise, challenge, redirect, or otherwise intervene in an AI decision. In situations where AI is making a fast decision, for example when AI powers a physical machine like a nuclear reactor, it will be even more difficult for a human to effectively "correct" the decision. For consumer AI, such as self-driving cars, the human truly becomes the "moral crumple zone," where the human is both subjected to potential harm and responsible for potential liability in their failure to intervene.[251] It is not consistent with principles of fairness to permit a liability shift here when the receiving party cannot adequately

---

[250] For example, limitations on liability are often barred in construction and engineering contracts. Buck S. Beltzer & Melissa A. Orien, *Are Courts Limiting Design Professionals' Ability to Limit Liability?*, 30 CONSTRUCTION L. 17, 17-18 (2010).

[251] *See* Elish, *supra* note 224, at 40.

perform the role consistent with a full transition of responsibility.

Second, AI developers may not consider how humans are included in the loop or required to work with AI during the AI design. More frequently, humans will be included as a "fail safe" for AI uses that seem particularly dangerous or could exhibit bias. Human involvement may be unavoidable, such as humans interacting with their medical devices or legally mandated secondary review of automated decisions, and in some cases, with appropriate training and expertise, humans could enhance AI function. However, for emerging AI technologies with unknown risk profiles, organizations may exploit humans in the loop primarily to offset liability risk. The underlying presumption is that humans can perform this role effectively, which, depending on the role they are expected to play, may not be a fair expectation. In this case, a valid liability shift should be premised on an underlying fact pattern consistent with a shift in responsibility based on the technology's design. While in some cases, a human may fully control the AI and have the knowledge and training to effectively do so, in most cases, humans will not have the requisite information available to factually accept risk from the AI creator.

Finally, a human in the loop may obscure real AI problems. Because AI development companies producing unknown or unsafe AI may exploit humans for their perceived ability to prevent undesirable outcomes (despite their inability to effectively do so), they have less incentive to test and monitor AI safety, efficacy, and fairness over time. If courts permit AI development companies to transfer liability to humans in the loop, these companies will not be held responsible for developing safe, efficacious, or fair AI. Under a new approach, courts should not enforce contractual liability transfers related to AI when they involve a transfer to a human person rather than a sophisticated organization to maintain fairness. In all cases, these arguments could be premised either as substantive or procedural unconscionability or as against public policy.

Overall, courts could develop a fairer liability model by refining products liability, making its applicability consistent across the AI regardless of embodiment, and offer some relaxation of pre-discovery evidentiary standards given the limitations in information and inscrutability of AI systems. Moreover, courts should consider barring the defensive application of limitations on liability for individuals positioned as liability shields while functioning as humans inside the loop.

Although the above model could improve the fairness of AI litigation *ex post*, the above model does not create any expectations for basic foundational design expectations or any other guarantees of reasonable behavior. For this reason, the United States should also consider a basic regulatory framework with a products liability reinforcing model.

III. A Humans Outside the Loop Liability and Regulatory Model

In Part I, this Article discussed the real practices that create the loop and the practices that differentiate between good and bad AI. In Part II, this Article evaluated tort liability models incorporating humans in the loop, which were revealed to frustrate their respective goals, create confusion, and potentially create systems of unfairness. These limitations on potential tort recovery *might* be worth it, if the collective system would lead to better AI and fewer injuries. Unfortunately, the degree to which these practices ultimately relieve AI development companies of liability leaves victims with no compensation and a cumulative market of unsafe products. Although the tort system can be better designed for AI harms, the cumulative harm that can occur in a universe of internet-tethered, AI directed products and the degree of that harm justify a regulatory approach that combines *ex ante* and *ex post* legal interventions.

This Part briefly introduces potential features of a regulatory model that reinforces a human-centered, responsible approach to ensuring AI is safe, accurate, and

fair.[252] This proposal includes the contours of a regulatory-tort model, which introduces *ex ante* regulatory requirements leveraging the advantages of administrative law coupled with a more consistent, predictable *ex post* products liability model. While this approach could appear to be heavy handed on its face, *ex ante* requirements could be drafted more generally to permit a degree of interpretation on the part of AI creators, flexibility to sector-specific expectations, and evolution as AI industry standards change.

A. Previous Proposals Should Inform Future Regulation

There are a wide variety of proposals that have emerged in the past five years, shaped by different presumptions about AI and their application in specific sectors.[253]

Ryan Abbott has proposed a model leveraging negligence law for human actors and products liability for non-human actors.[254] Abbott acknowledges that courts evaluate a human's unreasonable conduct under a negligence standard but that in limited circumstances strict liability may be used, as in products

---

[252] These features are introduced conceptually with the intent of offering a starting point for discussions about regulation. I will explore them in more detail in a future article.

[253] For example, much attention has been focused on the healthcare sector, which includes differing administrative behaviors and statutory limitations. *See, e.g.*, W. Nicholson Price II, Sara Gerke & Glenn I. Cohen, *Liability for Use of Artificial Intelligence in Medicine*, *in* RESEARCH HANDBOOK ON HEALTH, AI AND THE LAW (Barry Solaiman & Glen I. Cohen eds. 2023) (describing malpractice, derivative liability, direct hospital liability, and developer liability in healthcare); Frank Griffin, *Artificial Intelligence and Liability in Healthcare*, 31 HEALTH MATRIX 65 (2021) (examining a variety of liability models relevant to the healthcare sector, including products liability); Mindy Duffourc & Sara Gerke, *Decoding U.S. Tort Liability in Healthcare's Black-Box Era: Lessons from the EU*, 27 STAN. TECH. L. REV. 1 (2024) (comparing liability models for healthcare in the US and EU); W. Nicholson Price II & Glenn I. Cohen, *Locating Liability for Medical AI*, 73 DEPAUL L. REV. (forthcoming 2024) (positioning enterprise liability as an alternative to traditional tort liability doctrine); Tschider, *supra* note 111 (arguing for a reinforcement model involving regulation and torts when regulation does not prevent injury).

[254] Abbott, *supra* note 180, at 22-23.

liability.[255] Crucial to Abbott's proposal for AI tort claims is the impact on the use of AI, where treating AI differently from human actors in the same position (e.g., driving a car) could affect investment in automation.[256] Abbott reasons that AI operating in place of humans should therefore be held to a negligence standard.[257] However, such a negligence standard may also require interrogation of AI design, similar to how products liability design defect inquiries operate.[258] Finally, Abbott acknowledges that when AI cannot be held accountable as a tortfeasor, a human or organization will likely be responsible.[259] But human owners may not want to purchase AI if they will be held legally responsible, especially if insurance payments are high, and holding manufacturers responsible could stifle innovation.[260] Ultimately, Abbott acknowledges that manufacturers are in the best position to improve safety outcomes and that state limits on jury awards could avoid impacts on automation innovation, so long as a negligence standard, rather than a strict liability standard, is used.[261]

Despite the potential benefits of applying the negligence, there are also a variety of challenges. As Andrew Selbst describes, AI presents many challenges for negligence law.[262] Specifically, AI's "incomprehensibility" problem (described here as unintelligibility and inscrutability), "extreme corporate secrecy" (described here as legal opacity), and "replacement of individualized decision-making with statistical reasoning" significantly frustrate how negligence law functions.[263] To ameliorate some of these limitations, Selbst observes key differences that could be woven into negligence law's ambits. First, AI transforms established negligence concepts of

---

[255] *Id.* at 12-14.
[256] For example, autonomous driving could reduce the number of accidents and other injuries.
[257] *Id.* at 22.
[258] Abbott discusses an evidentiary standard that involves examining how "automation reduces, or is expected to reduce, *overall* accidents, not whether it did in fact reduce accidents in a specific instance." *Id.* at 27.
[259] *Id.* at 30-31.
[260] *Id.*
[261] *Id.*
[262] *See* Selbst, *supra* note 118, at 1360.
[263] *Id.*

foreseeability and bounded rationality, cornerstones of negligence, by creating scenarios that may be substantially more remote and inconsistent with the limitations courts usually place on this analysis.[264] Selbst argues that with available and intelligible explanations or interpretations, it will be possible to examine whether AI's results were foreseeable, even if "foreseeable cases will be the exception."[265]

Crootof, Kaminski, and Price recommend a model centered on the human in the loop, offering regulatory considerations for inclusion in a future framework.[266] First, they call for special consideration of the humans in the loop, particularly what roles they will occupy and when the law will mandate those roles.[267] This consideration is extremely important, as humans should not be placed in roles that they are unable to fulfill.[268] Special considerations of humans in the loop also inform potential system affordances.[269] The high degree of differentiation and consideration of context means that legal models will need to be flexible to accommodate these differences. "[G]eneralities . . . are at best little more than platitudes; at worst, they risk becoming influential but normatively problematic rules."[270] This context also informs where human intervention is desirable and why.[271]

Each of these approaches holds promise for creating consistency and promoting fairness for AI's human users. However, these approaches may not directly improve AI product safety overall. As explained more fully later in this Part, I recommend a combined regulation-tort approach that addresses the key issues raised by these scholars. This approach

---

[264] *Id.* at 1361.
[265] *Id.* at 1362. Selbst also notes that tort law could be effectively "sped up" in better establishing new standards for reasonableness through obligation or incentivization of "disclosure and pooling of knowledge about the common types of AI failures." *Id.* at 1368.
[266] *See* Crootof et al., *supra* note 1, at 487.
[267] *Id.*
[268] *Id.* at 489.
[269] *Id.* at 488.
[270] *Id.*
[271] *Id.* at 491-92.

avoids strict liability but bolsters a products-liability-based, design-defect model for torts that reduces both problematic negligence foreseeability issues and difficult-to-argue standards for products liability reasonable alternative design.[272] An adapted design-defect model will render limitations on liability, risk transfer, and user manual directions that are inconsistent with labels and marketing materials unenforceable. Additionally, regulatory requirements will be used to establish baseline standards of care, at least for AI development processes, to create shared expectations. Within such a regulatory model, AI must be designed to a safety standard which assumes no human will be available to supervise it. However, AI development companies may "design-in" humans in the loop and overcome this assumption by ensuring humans are adequately prepared to use, supervise, intervene in, or collaborate with AI. In this new model, an AI development company must demonstrate they have incorporated regulatory requirements into their product design, and a plaintiff's obligation to demonstrate reasonable alternative design is relaxed. Other entities, such as AI licensees, may also be sued, but the regulatory requirements they are required to follow will necessarily be different than AI development companies.[273]

---

[272] As previously noted, negligence standards and design-defect standards are markedly different. Under a negligence standard, plaintiffs must demonstrate that the defendant breached their reasonable duty of care by arguing what the reasonable duty of care was at the time. However, under a traditional products-liability design-defect model, the plaintiff is required to argue a reasonable alternative design that reduces potential harms with the same attendant benefits. The former standard is open-ended and informed by precedent, while the latter standard is highly specific and likely to fail based on lack of expertise and access to confidential information. This proposal adopt the most useful parts of both approaches to acknowledge information asymmetries but leverage the usefulness of a products-liability tort model.

[273] This Article proposes a model that includes both regulation and private right of action. However, I acknowledge that this may not be politically possible. In the event a regulatory approach is not possible, courts may also consider adopting a products liability model for AI products, even if they are not embodied, and permitting some flexibility for plaintiffs in arguing a reasonable alternative design for the reasons explored in this Article.

B. A Regulatory-Tort Model Improves Consistency and Promotes Fairness

Often, regulatory approaches and the common-law tort system are positioned as alternatives to one another, with the administrative state posing some risk to the tort system.[274] The common-law system is often a poor choice to prevent harm resulting from emerging technology due to the niche knowledge implicit in technology development. By contrast, regulatory agencies have a much stronger ability to develop this expertise, allocate risks and benefits across a population, engage in rule-making, and adapt quickly to changing conditions.[275] However, this perception of administrative agencies may differ from reality: "The appealing image of a centrally located, expert agency, conducting a balanced and detailed quantitative analysis, overstates the practical abilities of regulatory agencies . . . . Agencies do not have a monopoly on expert knowledge."[276]

Both agencies and courts rely on experts to advise on legal issues.[277] In some cases, courts may have the time to examine issues more thoroughly, through a learning process.[278] Tort law's temporality can also be an advantage because it allows courts to capture evolutions in social norms.[279] Tort also has an important compensatory function, aiming to compensate plaintiffs for their injuries.[280] Regulation generally does not serve a reliable compensatory role, even if an agency fines

---

[274] *See* Catherine M. Sharkey, *The Administrative State and the Common Law: Regulatory Substitutes or Complements?*, 6 EMORY L.J. 1705, 1729 (2016) (describing Justice Scalia's view that state torts are a poor substitute for the FDA's *ex ante* rigorous pre-market approval process).
[275] *See* Mary L. Lyndon, *Tort Law and Technology*, 12 YALE J. REGUL. 137, 153 (1995).
[276] *Id.* at 154-56.
[277] *Id.* at 158.
[278] *Id.* at 162-64.
[279] *Id.* at 163-64.
[280] *See* Tschider, *supra* note 111, at 1614 (describing the need for compensation for patients harmed by medical AI despite regulatory review and express preemption language).

certain organizations and establishes a civil penalty fund.[281] There is one feature, however, that the common law cannot adequately provide: preventing harm from occurring *ex ante*.[282] At its foundation, tort law is responsive, focused on allocation of costs, distributive justice, and compensation—but only after harm has occurred. Regulation has the benefit of preventing at least some injuries from occurring, at least for organizations complying with the law, though likely not all of them. For emerging technologies, only so much can be known at the time of regulation drafting, and these technologies may undergo substantial change in a relatively short period of time.

Further, tort law can have deterrence capabilities, through litigation generally and punitive damages in particular. The challenge is that common-law deterrence can move very slowly and by accretion. In contrast, once an effective date is selected, direct regulation moves quickly. In the event administrative agencies are permitted by law to promulgate rules and draft guidance, these activities can create greater specificity and respond to changing standards, while negotiating rules between agencies when needed. For certain industries, agencies do not just enforce the law, they can perform a gate-keeping role, preventing overly dangerous technologies from being available.[283]

Tort and regulation advance different goals, but they can also be useful in reinforcing each other's goals. In a regulatory-

---

[281] *Payments to Harmed Consumers*, U.S. CONSUMER FIN. PROT. BUREAU, https://www.consumerfinance.gov/enforcement/payments-harmed-consumers [https://perma.cc/95DQ-6CXQ].

[282] This may be why some scholars suggest licensing or other authorization schemes be put in place prior to AI being available for use. *See* Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 394 (2016) (proposing an agency for certifying AI). I reject a certification mechanism, as it would likely slow down innovation and substantially increase costs for primarily start-ups and small businesses in this important space. But it may be beneficial for certain sectors to adopt regulatory approval approaches for greater risk areas, such as medical device regulation.

[283] For example, the FDA arguably keeps overly dangerous medical devices and pharmaceuticals out of the marketplace under most circumstances by ensuring products are reviewed before being made available to the public. *See* Sharkey, *supra* note 274, at 1730-32 (describing various Supreme Court decisions examining the administrative-tort regulatory trade-off involving medical devices and pharmaceuticals).

tort model, tort law performs a reinforcing role by enabling enforcement of statutory provisions by harmed plaintiffs. Additionally, administrative agencies may have the ability to prosecute for non-compliance or bar product sales through injunctive relief or revocation of clearance to sell. The importance of a regulatory model is to create *ex ante* general guardrails for AI's function that can evolve over time. One reasonably successful approach is to develop statutory requirements that are process and capability-based, while leveraging organizational expertise to tailor these processes to an organization's practices.[284]

David Thaw calls this Management-Based Regulatory Delegation, an approach where "administrative agencies, through legislative-mandated collaboration with regulatory stakeholders, promulgate regulations requiring regulated entities to develop plans designed to achieve certain aspirational goals laid out by the legislature."[285] One advantage to Management-Based Regulatory Delegation is that organizations can meet statutory requirements a variety of ways, promoting innovation not only in product development but also in compliance. This approach is evidenced in the Health Insurance Portability and Accountability Act and the Gramm-Leach Bliley Act, two significant federal laws that incorporate privacy and security requirements.[286] Crootof, Kaminski, and Price analyze existing U.S. regulation of three "safety critical systems," railroads, nuclear power, and medical devices: all three frameworks incorporate some combination of training requirements, licensing regimes, and guidance.[287]

---

[284] This form of regulation leverages the knowledge of regulated entities that may actually have more expertise than the regulator, while also offering regulators some discretion in enforcing the regulation.

[285] David Thaw, *The Efficacy of Cybersecurity Regulation*, 30 GA. ST. L. REV. 289, 293 & n. 18 (2014).

[286] *Id.* at 296.

[287] Crootof et al., *supra* note 1, at 492-97. Crootof, Kaminski, and Price also examine recent developments in the European Union. *See id.* at 503-04. Although certainly much can be gleaned from the European Union's globally leading efforts in this space, this Article examines the problem from

Although these models differ, they all offer the advantage of expert, *ex ante*, involvement to promote safety. While the tort system alone could be adjusted to apportion liability more fairly for emerging technologies like AI, a reinforcing *ex ante* approach prevents harm, at least to some degree, while creating consistency in legal expectations for organizations.

C. A Statutory Approach Can Eliminate Human Liability Shields

A regulatory approach must be designed to benefit humans, rather than to exploit them.[288] Key to this model is improving safety, efficacy (or accuracy), and fairness in AI. Safety and efficacy mean that the risk and reward are adequately balanced, that AI does not cause unreasonable risk to humans and property (safety), and AI works effectively for its designed purpose (efficacy). Fairness examines how AI performs for a given individual in society. For example, if AI is reasonably safe for one community but unsafe for individuals within another, it is likely to be unfair and even potentially discriminatory, based on its effects. When AI is not embodied, treating humans unfairly could also result in denial of benefits or opportunities for certain groups of people. By designing regulation around potential risk to humans, AI regulation will necessarily incorporate responsible AI standards.[289] To ensure reasonably safe, effective, and fair AI reaches the marketplace, it will be essential to select a regulatory authority that statutorily promulgates rules, regulations, and guidance, which can update as AI evolves.

Although previous proposals would undoubtedly improve the current state of affairs, a regulatory-tort approach establishes what a solely *ex post* recovery model cannot: at least for some people, it will prevent harm instead of only permitting

---

a U.S.-centric perspective. For more examination of the European Union's progress in AI regulation, see Margot Kaminski, *Regulating the Risks of AI*, 103 B.U. L. REV. 1347, 1372-88 (2023) (describing how the E.U.'s regulation has oriented around AI risk, including how to determine potential impact).

[288] For a discussion of ways the law views exploitation as harm, see Tschider, *supra* note 74, at 888.

[289] *See* Karen Yeung et al., *AI Governance by Human Rights-Centered Design, Deliberation, and Oversight*, *in* THE OXFORD HANDBOOK OF ETHICS OF AI 76, 88-93 (2020).

injured parties the possibility of recovery. Additionally, the presence of one or more regulatory bodies ensures that, at least to some degree, investigations responding to valid complaints can take place before anyone (or at least a larger number of people) is injured. These investigations may lead to required remediation or, in extreme cases, removal from the marketplace. The consistency of rules also should improve trust within the AI supply chain, generating trust between AI creators and licensees and between users and manufacturers.

It is extremely important for AI, some of which may present significant risk to humanity, to adhere to consistent processes, be subject to periodic audits, and be overseen sufficiently by some regulatory body. However, it is inevitable, like for any consumer products, that some individuals will be injured as a result of AI use. In these cases, a private right of action, limited to claims associated with statutory obligations, should be made available as a reinforcing mechanism *ex post*.

1. Regulatory Authority

An initial question surrounds which regulatory agency, if any, is positioned to regulate AI. Because individuals and businesses can use AI across a number of sectors, consolidating AI regulation runs at cross-purposes with the United States's dominant sectoral regulatory approach.[290] A subsidiary question of who should regulate is whether general or sectoral regulatory activity is more effective in improving organizational behavior. If a new agency were to regulate AI, it could create challenges for administrative agencies that already actively review or regulate such technology, like the Federal Trade Commission (FTC) (regulating consumer products), the National Highway Safety Commission (regulating autonomous vehicles), or the FDA (regulating AI

[290] Jacob Turner observes that narrow AI will eventually shift to general AI. Developing a regulation is important for this shift. Moreover, general principles are likely to be more successful than industry-oriented ones. *See* TURNER, *supra* note 152, at 213-19.

medical devices).[291] However, there are many examples where regulatory agencies work in concert, coordinating regulatory activities.[292] For example, while the FDA conducts post-market surveillance of devices and pharmaceuticals and is primarily responsible for labeling claims, the FTC is responsible for advertising claims and may prosecute false or misleading labeling claims on the pharmaceuticals.[293] It is necessary for the FTC to work with the FDA to understand the context of labeling claims, for example medical efficacy claims and potential side effects.

Although the FTC has emerged as the *de facto* privacy regulator and has begun regulating AI under the unfairness prong of Section 5 of the FTC Act, the FTC's ability to engage in rule-making is somewhat limited by the Magnusson-Moss Warranty Act.[294] While the FTC is uniquely positioned to

---

[291] Andrew Smith, *Using Artificial Intelligence and Algorithms*, FED. TRADE COMM'N (Apr. 8, 2020), https://www.ftc.gov/business-guidance/blog/2020/04/using-artificial-intelligence-and-algorithms [https://perma.cc/T6QM-3CC9]; Occupant Protection for Vehicles With Automated Driving Systems, 49 C.F.R. pt. 571 (2021); *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan*, U.S. FOOD & DRUG ADMIN. (Jan. 2021), https://www.fda.gov/media/145022/download [https://perma.cc/9V8G-MY4L].

[292] *See* Guihot et al., *supra* note 226, at 424-25.

[293] Annie M. Ugurlayan, *FTC's New Health Claims Guidance: What You Need to Know*, BBB NATIONAL PROGRAMS (Feb. 15, 2023), https://bbbprograms.org/media-center/bd/insights/2023/02/15/ftc-health-claims-guidance [https://perma.cc/Q7NJ-RV8B]; *How FDA Approves Drugs and Regulates Their Safety and Effectiveness*, CONG. RSCH. SERV. 15-17 (May 8, 2018), https://crsreports.congress.gov/product/pdf/R/R41983 [https://perma.cc/A77S-HASM]; The Fair Packaging and Labeling Act of 1967, 16 C.F.R. pt. 500 (directing the FTC and FDA to issue labeling regulations for all consumer commodities).

[294] Kelley Drye, *The FTC's Magnuson-Moss Rulemaking Process—Still An Uphill Climb*, KELLEY DRYE (Jan. 10, 2022), https://www.kelleydrye.com/viewpoints/blogs/ad-law-access/the-ftcs-magnuson-moss-rulemaking-process-still-an-uphill-climb [https://perma.cc/JSH6-NVUL]. The FTC has legal authority under the Federal Trade Commission Act to enforce against unfair or deceptive trade practices. *See generally* Federal Trade Commission Act, Pub. L. No. 447, ch. 49, § 3, 52 Stat. 111, 111 (1938) (proscribing "unfair or deceptive acts or practices" and "unfair methods of competition"). Predominantly, the FTC has enforced against deceptive trade practices. However, when companies

regulate on the basis of consumer protection, many AI are sold to companies, not consumers, and many resulting efficiency gains are estimated to occur within organizations.[295] The Consumer Product Safety Commission (CPSC) is another example of a consumer-oriented administrative agency that could possibly regulate AI. The agency is authorized to develop standards, recall products, and ban the sale of certain products.[296] There are a variety of other agencies that could regulate AI, at least within their sector. However, the challenge in naming an existing regulatory authority is that their current scope of enforcement and ability to effectuate rulemaking is set, at least without additional updates to the law.[297]

The cleanest way to implement an AI regulation is to create a new agency, for purposes of this Article called the "U.S. Artificial Intelligence Administration" (AIA), specifically focused on AI regulation across sectors with the statutory obligation to adopt AI rules and guidance.[298] The United States has experience conducting research, interviewing industry, and establishing technology standards. The National Institute of Standards and Technology (NIST) has led many recent

---

have not lied but simply failed to perform as expected, as could occur in an AI malfunction or an undesirable data breach, the FTC has struggled to argue that these are unfair business practices. *See, e.g.*, LabMD v. FTC, 894 F.3d 1221, 1231 (11th Cir. 2018). The FTC has been proposed as a potential regulatory agency for AI. *See* Michael Spiro, *The FTC and AI Governance: A Regulatory Proposal*, 10 SEATTLE J. TECH., ENV'T & INNOVATION L. 26, 52-54 (2020).

[295] *See* Katherine Haan, *How Businesses Are Using Artificial Intelligence in 2024*, FORBES ADVISOR (Apr. 24, 2023, 7:54 AM), https://www.forbes.com/advisor/business/software/ai-in-business [https://perma.cc/ZTT8-GS2H]. Almost 97% of businesses believe ChatGPT, for example, will help their business, and 64% believe AI will improve their relationships with customers. *Id.*

[296] *Regulations, Laws & Standards*, U.S. CONSUMER PROD. SAFETY COMM'N., https://www.cpsc.gov/Regulations-Laws--Standards [https://perma.cc/E4KE-FBFF].

[297] *See* Tschider, *supra* note 93, at 125 (describing the FTC's rulemaking challenges after the passage of the Magnusson-Moss Warranty Act).

[298] *See* Guihot et al., *supra* note 226, at 443-44 (advocating for flexibility in rulemaking, ability to engage in anticipatory rulemaking, and collecting data for purposes of data-informed decision-making).

developments in cybersecurity risk management and has begun to issue similar risk management processes for AI.[299] However, NIST is not a regulator and has no ability to enforce anything. The AIA could benefit from the extensive work completed to date, for example to address looming AI cybersecurity concerns,[300] and convert the most useful portions of NIST's AI risk management work into binding rules that the AIA can enforce, where needed. In the event products disproportionately affect a consumer base, the AIA could work with the CPSC and FTC, depending on the factual circumstances, to collectively investigate and prosecute. For certain highly developed regulatory regimes, like the FDA, the AIA could consult but leave enforcement to the FDA.

Because AI products are very diverse in functionality, users, and potential risk, the AIA would not be well-positioned to provide broad clearance or approval for sale.[301] Rather, the AIA could spend more time investigating products and enforcing violations of AI standards. The AIA would develop its expertise over time and use discretion to enforce against higher risk entities.[302] In sectors where clearance or approval is already required by an administrative body *ex ante*, such as for medical devices, the administration for that sector would provide approval with the option of including experts from the AIA in meetings concerning AI's safety, efficacy, and fairness.[303]

Although the AIA could perform investigations of AI developer or licensee practices, the AIA will not likely receive the staff to routinely audit AI development practices or subsequent licensing. However, similar to audits conducted by

---

[299] NATL. INST. STDS. TECH., *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, NIST AI 100-1 (Jan. 2023), https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [https://perma.cc/F2BW-N94D].

[300] *See* Tschider, *supra* note 93, at 97-98.

[301] For example, one criticism of the FDA is that they are unable to understand the detailed functionality of medical devices reviewed in part because of their relative diversity and sheer volume. I expect a similar challenge would apply to the AIA if the agency was involved in AI clearance.

[302] *See* Guihot et al., *supra* note 226, at 450-51.

[303] This could help supplement expertise differences between the AIA and these entities.

the Office of the Comptroller of Currency, the AIA could select organizations at random or review them on a schedule regarding their AI development practices (scheduled by the level of inherent risk in the products or by the gross profits of the firm). In this case, a centralized assessment or audit function, such as the former (currently inactive) U.S. Office of Technology Assessment, could perform an important role auditing technology practices. Conducting periodic audits of AI product development practices can help to identify potential issues and ensure they are corrected on a timely basis, without solely relying on outside complaints and investigations.[304] In addition, conducting audits across a population of AI developers would likely reveal universal AI challenges and imperfections in the AI statute, leading to informed updates to associated rules and guidance.

To ensure mandatory internal governance processes occur, the AIA or associated rules and legislation could require board reviews and internal approvals. For example, a new, foundational federal AI statute could require the instatement of Institutional Review Boards (IRBs) or mandatory ethics boards, common in the life sciences and research and development fields, to be operated by the organization creating AI.[305] The boards would be responsible for evaluating

---

[304] Olga V. Mack & Emili Budell-Rhodes, *Navigating the AI Audit: A Comprehensive Guide to Best Practices*, LAW.COM (Oct. 20, 2023, 12:15 PM), https://www.law.com/legaltechnews/2023/10/20/navigating-the-ai-audit-a-comprehensive-guide-to-best-practices [https://perma.cc/Y5N8-4LQF]. One option for conducting centralized auditing or validating self-assessment is through the now-retired Office of Technology Assessment. *The Office of Technology Assessment*, GOV'T ACCOUNTABILITY OFF. (Oct. 13, 1977), https://www.gao.gov/products/103962 [https://perma.cc/X2UG-BN2A].

[305] Reid Blackman, *Why You Need an AI Ethics Committee*, HARV. BUS. REV. (Jul-Aug 2022), https://hbr.org/2022/07/why-you-need-an-ai-ethics-committee [https://perma.cc/D3E5-YV99]. Under the Common Rule, all organizations that are conducting clinical testing on humans are required to have these activities reviewed by an Institutional Review Board. *See* Federal Policy for the Protection of Human Subjects (45 CFR 46). The degree of board review is generally informed by how risky the activity—

enterprise risk and AI priorities, as well as the AI's effects on humans. While such boards could be mandatory for AI development companies, they could also be mandatory for organizations using AI in their products or licensing AI to others. Where AI will be designed for human use, includes a human in the loop, or could affect a human physically, individuals who use the products could participate in a similar review board focusing on effective AI design for humans.[306] The human design board could be required to evaluate AI for its human-centered approach from testing to ongoing live use, aiming for minimal negative impact and maximum positive benefit to the human. In situations where a human in the loop is unavoidable, the IRB could evaluate the plans for ensuring minimal physical and psychological risk to the human as well as requiring adequate interface design and training, as needed. Because IRBs are not infallible, decisions could be reviewed by the AIA or another regulator periodically or if the AIA receives a complaint about AI practices.

2. Definitions

Definitions are essential to determine the obligations that apply to different organizations and the subject matter of the regulation. Given the information asymmetry and opacity between AI developers and users (who may be harmed), specific roles will differentiate obligations between consumers using AI and organizations that create it.[307]

The first definition would apply to AI development companies, which could be called "AI developers."[308] AI

---

some activities may be exempt from review, while others require a more comprehensive evaluation. *Id.* For example, pharmaceutical clinical trials and medical device clinical trials must be overseen by an IRB. An IRB's goals include minimizing risk to human subjects, balancing risks and benefits, and facilitating consent. Maria I. Lapid, Bart L. Clarke & R. Scott Wright, *Institutional Review Boards: What Clinician Researchers Need to Know*, 94(3) CONCISE REV. CLIN. 515, 516-17 (2019).

[306] Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1045-46 (2014).

[307] *See supra* Section II.B and Section III.B.3.

[308] I choose this language because commentators have frequently called for AI developer liability, which presumes an individual can be held accountable. But AI will generally not be created in a way that would allow

developers, as the creators of the vast majority of AI features and functions, will be the most heavily regulated and will be expected to perform any number of statutory duties to create AI that is safe, accurate, and fair, at least with respect to the potential benefits AI can offer. Failure to perform these statutory duties of care will likely result in liability, if the failures proximately caused a plaintiff's injury. Part of the challenge in regulating developers is that developers may not know how their AI models will be used in downstream implementations. However, they are best positioned to communicate, train, and label AI with respect to its designated use and to ensure human outside the loop efforts (in design and testing) are done adequately with respect to its expected use. Moreover, the AI development companies' products may disclose their intended use.

However, organizations that acquire, license, and use AI should also be regulated since they are using AI in ways that could affect other organizations, people, and property. There are some organizations that have less information but have the power and opportunity to legally obtain it. For example, AI acquirers, organizations that legally acquire AI through extensive due diligence processes, should be held to the same obligations as the original development company, which likely no longer exists after acquisition. Businesses acquiring technology already engage in due diligence activities involving potential risks of acquisition, though arguably acquirers may not have historically spent as much time examining the technology itself.[309] By holding acquirers to the same standard, regulators will require acquirers to examine the AI they wish to acquire.

---

it to be released into the market without the backing of a company. For this reason, "AI developer" can be used but should be defined as an *organization* developing AI products.

[309] Imran Ahmad, Roxanne Caron & Suzie Suliman, *Deal-Maker or Deal-Breaker: The Legal Ins and Outs of Using AI in M&A*, LEXPERT (July 10, 2023), https://www.lexpert.ca/legal-insights/deal-maker-or-deal-breaker-the-legal-ins-and-outs-of-using-ai-in-ma/377272 [https://perma.cc/SM8K-XZB3].

Second, organizations that license AI, "AI licensees," may be challenged in what information they can obtain about AI developers and therefore should not be responsible for every harm resulting from AI. However, licensees should be required to perform reasonable third-party assessments to ascertain what risks AI introduces and document their findings. AI licensees then will be responsible for any risks (and associated harms resulting from realized risks) they accept as part of using the product or integrating it into a technology stack. Although an AI licensee could sue an AI developer for breach of contract or implead to share or accept liability, they will be responsible for the risks they knew about and accepted. If an AI licensee does not conduct a third-party assessment, it accepts all potential risks as a type of statutory liability transfer. The goal in this approach is not to penalize licensees for unforeseeable harms introduced by AI developers that would be undiscoverable through reasonable inspection or testing; however, licensees must also take care in selecting AI development companies and reviewing the AI to the extent it is possible.

Finally, AI will be used by people who may ultimately be harmed by it. Consumers, or individual AI users will not be held accountable under the proposed AI statute for negligent action or inaction related to AI when it causes harm. However, users who engage in reckless or intentional behavior (i.e., behavior outside of designed AI functionality or expected human actions in relation to it) that causes harm in conjunction with an AI product may be held accountable under traditional torts. For example, drivers using an autonomous vehicle could be responsible for harms caused when their autonomous vehicles have exhibited major issues over time, and the driver does nothing. Or, a human could be legally responsible when they know how an AI system will malfunction and they knowingly use this to cause harm. These types of torts should dovetail nicely with existing torts because the AI is only an instrument of a human plaintiff's actions. In addition to definitions concerning the technology and various roles in the AI supply chain, a regulatory solution could prime courts to apply products liability to all AI based on its definition as a "product." Defining AI as a product clarifies the body of law that should apply, creating consistency across various forms of

AI, whether they are embodied or not. I will explore this concept in more detail related to a private right of action.

It will be crucial to determine how organizations might be exempted from this statute. A common exemption could be based on organizational size, operating income, or revenue. In this case, however, organizations creating AI may be start-ups and small businesses, some of which may no longer be in operation long after the AI is created and sold. For this reason, it is important to ensure that AI is created responsibly, even if an organization is small. Another important consideration is whether a proposed AI statute should be applicable across all sectors, similar to President Biden's Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI which has directed all administrative agencies to examine their AI policies.[310] Although a new AI statute could serve as a floor for all AI organizations, some regulatory approaches could be exempted, so long as they can at least meet the minimum requirements established under the statute. However, broadly exempting entire sectors could also result in lesser standards established.

3. Humans Outside the Loop: Design, Testing, and Governance

To develop objectively safe and fair AI, AI developers should center AI design, training, and testing processes around the expected uses, sectors, and applications of AI. To offer flexibility, the proposed AI statute should establish baseline regulatory standards of care for AI developers, acquirers, and licensees. The most important focus of a regulatory standard of care is establishing standards for data selection and model development, which should be oriented towards expected sectors, data types, users, and functional applications. For example, a safe and accurate AI design for medical devices is likely quite different from a safe and effective design for GAI applications to automate customer service chats. For this reason, the AI statute should reference regulatory requirements, while the AIA, NIST, and other agencies

[310] Exec. Order No. 14110 (Oct. 30, 2023).

(depending on the field) could establish specific rules, guidance, and standards. In administrative rules and guidance, general regulatory requirements can become more specific; this regulatory approach also enables faster updates in response to changing technology.[311] In all cases, AI design should anticipate potential issues that could arise in testing. For example, the potential for discriminatory or unfair function must be anticipated and planned for based on the selected learning processes and the data sets used to develop AI. The statute should require all design documents to be retained for at least ten years after the AI is released, to be available in investigations or litigation, including designs not selected and design changes made.

Crucial to functional AI design is the user's experience as they interact with the AI. The Human-Computer-Interaction field has evolved over many years to refine practices that optimize human experiences with computers. For AI that includes a user interface or AI that requires a human in the loop, the AI statute should require specific consideration and design of the human interface, building on a division of Human-Computer-Interaction known as human-computer collaboration (HCC).[312] HCC is defined as "a process in which two or more agents work together to achieve shared goals," which "involv[es] at least one human and one computational agent."[313] When HCC is being used, AI development companies must design for the human-computer interface, including identifying potential malfunctions and communicating, training, or designing to address these

---

[311] The recent executive order directs NIST to establish key benchmarks for AI. *See id.* As a standard-bearing body, rather than a regulator, NIST's benchmarks and standards will provide useful direction for regulators charged with directing AI regulation within their regulatory arenas. For this reason, NIST should be part of conversations involving guidance and rules to interpret an AI statute for different AI sectors.

[312] *See* Terveen, *supra* note 105, at 67. This approach could be consistent with other calls for licensing developers or humans in the loop. *See* Chinmayi Sharma & Benjamin C. Zipursky, *Who's Afraid of Products Liability? Cybersecurity and the Defect Model*, LAWFARE (Oct. 19, 2023, 10:24 AM), https://www.lawfaremedia.org/article/who-s-afraid-of-products-liability-cybersecurity-and-the-defect-model [https://perma.cc/2NAH-UD4Z].

[313] *See* Terveen, *supra* note 105, at 67.

malfunctions. While not every malfunction is foreseeable, AI development companies will have satisfied their duty of care when they can validate that they have reasonably designed such interfaces to enable a human to identify potential high-risk malfunctions and avoid harm.

As a starting point, to avoid exploitation of humans in the loop, the AI statute should prohibit companies from considering humans in the loop when assessing risk or designing appropriate safety measures *unless* an AI developer can demonstrate they have comprehensively designed the AI for human's use. This presumption against humans in the loop is not to outright bar human participation with AI but to ensure that humans are not used as an alternative to safe, efficacious, and fair design. Designing for humans in the loop optimizes their role by ensuring that fewer human errors occur and by enabling human users to spot situations where AI is not functioning as intended. There may be many valid situations where humans should be in the loop, but they may not be exploited to generate false trust in product safety or for purposes of shifting liability to them.

To design for HCC, systems must include allocation of responsibility and coordination. Specifically, systems must ensure that agents are not designed for the exact same task, but rather for different tasks within the system.[314] Agents, both the computer and the human(s), must also be aware of and able to gauge where the other agent is in the process.[315] Special attention must be given to communication and anticipation of potential problems that can affect one agent's ability to fulfill its task, and agents must be positioned to adapt and learn, changing the AI model to fulfill the shared goals.[316] AI developers and licensees using AI in their technology stack to interact directly or to collaborate with humans must be required to evaluate their practices using HCC concepts or with an HCC professional. Any decisions regarding design and

[314] *Id.*
[315] *Id.*
[316] *Id.*

function pertaining to the human experience must be reported to the IRB and human review board for evaluation.

Following AI design, AI creators must focus on the process of training algorithms for the AI's expected uses. Training can be quite variable depending on the type of AI being used and the method of its use. For example, for machine learning, neural networking, and deep learning approaches, learning types may be supervised or unsupervised. The training approach is directly tied to design of the system overall, which means it will be difficult to create a statute with specific required training activities that meet a standard of care. In some cases, training should be accompanied by expert input and evaluation or verification steps, especially when the system is designed to replicate an existing complex process currently performed by humans (such as medical diagnostics). A requirement to reasonably engage in responsible training with respect to AI's expected uses and users offers flexibility in the training process while also establishing an expectation for these practices.

Testing can also be designed for any variety of purposes. For example, software development testing practices, such as performance testing, User Acceptance Testing for systems with a user interface, and cybersecurity penetration testing should be used, and failure to perform these testing protocols at all would demonstrate a breach of the standard of care.[317] However, many types of AI-specific testing will be needed to ensure the live version of AI operates without a large number of issues. The AI statute itself does not need to mandate all types of testing, but it can minimally require testing practices be used to reasonably avoid downstream risks, and it can require specific types of testing like penetration testing or bias testing for all AI development.

Central to all decisions involving design and testing is internal organizational oversight. Governance is the process of collective decision-making about an organization, which includes not only leadership and decisional processes but also documentation and communication of practices to which an organization desires to hold itself accountable. In short, governance is *self*-governance, where an organization

[317] *See* Yeung et al., *supra* note 289, at 98.

establishes its own rules of operation. Governance practices usually begin with creating an internal organization that will ensure all internal activities occur. For smaller organizations, the internal organization could be shared, for example a combined role with privacy, cybersecurity, or quality management. For larger organizations, a specific department could be created to manage governance activities. To ensure some accountability occurs at the top, an AI statute should mandate the naming of an AI leader to oversee the program with specific accountability to AI safety, efficacy, and fairness.

Governance activities usually begin with policy development. Policies should, at a minimum, communicate both the requirements under the AI statute and how the organization (and all personnel as an extension of it) plans to meet these requirements. This includes specific policies, standards, and procedures that document how processes will be run within the organization. These documents must be provided to AI licensees upon request, but subject to confidentiality obligations, such as non-disclosure agreements.

The AI statute should require risk assessments to be described in any governance policy. Assessing risk is the practice of reviewing either your own practices or those of another entity, through the lens of particular kinds of risks, such as privacy risk, cybersecurity risk, safety hazards, or risk of bias.[318] Risk assessments evaluate risks viewed from any number of lenses; a risk assessment focused on design safety will examine AI differently than a cybersecurity risk assessment. The purpose of a risk assessment is to identify potential issues before they occur and to make informed decisions about identified risks, closing, remediating, or

---

[318] *What Is a Risk Assessment? (With Benefits, How-To and Types)*, INDEED (Apr. 11, 2023), https://www.indeed.com/career-advice/career-development/risk-assessment [https://perma.cc/99Q4-CA2D]; Ilana Golbin, *Algorithmic Impact Assessments: What Are They and Why Do You Need Them?*, PWC (Oct. 28, 2021), https://www.pwc.com/us/en/tech-effect/ai-analytics/algorithmic-impact-assessments.html [https://perma.cc/GP44-NXPS]; *7 Steps in Privacy Risk Management*, DATA PRIV. MANAGER (Nov. 5, 2022), https://dataprivacymanager.net/7-steps-and-elements-of-privacy-risk-management [https://perma.cc/4QCS-S4FT].

mitigating any risks or avoiding courses of action. Although risk assessments could be conducted by an external party on behalf of the AI developer or licensee,[319] these organizations will still need to engage in the governance process of making decisions about identified risks (and weighing the potential for legal action against them). Such decisions should be required to be documented under the AI statute and not shielded by privilege, which will make them subject to discovery proceedings.[320]

4. Humans Outside the Loop: Feedback and Reviews

Although much has been made of ChatGPT's inaccuracies and hallucinations, ChatGPT's story has demonstrated a version of success through rigorous public use and black box testing. Feedback loops are essential to AI development over time; learning what works, what fails, and how AI has not performed as expected, informs how AI systems should be changed and improved. These feedback loops based on human use of AI can reveal foundational AI design problems and user preferences. In contrast with humans in the loop, these humans outside the loop (after the loop) are essential to developing AI that improves over time. Simply providing feedback and building a mechanism for it within the tool or outside the tool is not enough, however.

In addition to feedback, internal monitoring and inspection practices are important, as is adverse testing. This testing, performed by entities adverse to an organization's interests

---

[319] AI practices could be reviewed and actually certified. *See* Virginia Dignum, *Responsibility and Artificial Intelligence*, *in* THE OXFORD HANDBOOK OF ETHICAL AI 214, 225 (2020). Although certification could be required as part of a statute, such a statute would likely presume that organizations capable of offering the relevant certifications already exist and that relevant legitimate and legal practices are well-known. Today, AI is still fairly experimental, so it could be difficult to mandate or otherwise offer incentives for certification at this time.

[320] Governance decisions cannot be shielded by privilege because they will be the best source of information related to the questions of foreseeability and alternative design. In the event that all of these documents are shielded by privilege, it will be nearly impossible for any plaintiff to recover, despite suffering harm resulting from design defects. *See* Daniel Schwarcz, Josephine Wolff & Daniel W. Woods, *How Privilege Undermines Cybersecurity*, 36 HARV. J.L. & TECH. 421, 478 (2023).

(such as competitors), occurs outside the loop. While it could be useful to have more extensive regulatory oversight, changing administrative priorities may underscore the importance of reviews that are not wholly dependent on administrative personnel. One approach to more comprehensive monitoring of AI outside the loop involves crowd-sourcing feedback. With the ability to inspect and evaluate practices, people (in the United States and abroad) feel more connected and engaged in government practices. A wide variety of skills will also enable the identification of potential issues faster and hopefully before anyone is impacted.

One key issue where the public could add tremendous value is in evaluating live AI. Some AI, such as those designed for a narrow purpose, are not used by enough people to generate sufficient feedback to fine-tune an AI model,[321] but exposing AI to a larger group by hosting it publicly could crowd-source finding potential weaknesses or vulnerabilities in the tool. One option is to mandate AI systems with sufficient complexity, such as those using neural networks, to host a version of their algorithms and user interface in a public forum so that any number of humans can identify potential issues and even provide feedback on the AI product.[322] AI systems that are less

---

[321] For AI to function optimally, feedback is needed during its use (outside the loop) to improve it over time. This fine tuning differs from broad-scale testing, in that the AI already is performing at some expected level, but fine tuning enables the AI to become more effective over time as recommendations and decisions are confirmed or rejected. Formally, fine tuning is where data scientists apply a broader mile to more distinct (narrow) use cases, but some of this work may be conducted in its live version. Lev Craig, *Fine-Tuning*, TECHTARGET (Feb. 2024), https://www.techtarget.com/searchEnterpriseAI/definition/fine-tuning [https://perma.cc/74QC-4KJZ].

[322] Norm enforcement and, ultimately, external governance results in part from public opinion. *See* Urs Gasser & Carolyn Schmitt, *The Role of Professional Norms in the Governance of Artificial Intelligence*, *in* THE OXFORD HANDBOOK OF ETHICAL AI 141, 154-55 (2020). For example, systems could undergo public "discrimination stress testing," where organizations are required to simulate their AI on populations prior to release in order to avoid discriminatory impact. *See* Talia B. Gillis & Jann

complex and could be reverse engineered could instead host a page for inquiries, reports, and feedback. In all cases, registration of AI (and a basic description of its designed uses, along with contact information) should be mandatory. President Biden's AI Executive Order contemplates this concept, directing the Secretary of Commerce to report on dual-use foundational models, or AI that is very large and general purpose.[323] The goal is to protect national security.[324] However, there are many AI that could harm people, property, and economic interests that would be less concerning from a national security perspective. Registering these systems more broadly will enable administrative agencies like the AIA and any other applicable sectoral agencies to review their practices.

There are a variety of positive outcomes in registering AI companies, hosting live AI for feedback, and requiring its registration. For example, security researchers might find and report bugs privately through a communication mechanism on the hosting site. AI fairness researchers might report malfunctions that disproportionately affect certain people. The general public might spot inaccuracies, or companies could host bug bounties to identify them. In all cases, issue spotting is something that humans do tremendously well in comparison to machines because it is inherently contextual. The United States might also discover types of AI that carry higher inherent risk, which could potentially mean certain technologies are temporarily barred from use, are subject to strict liability, or require insurance coverage.[325]

L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV. 459, 481, 484 (2019) (describing stress testing conducted prior to public release, though such testing could be conducted on a public version of AI, too). It could be possible to confirm these tests through crowdsourcing, as well.

[323] *See supra* note 310.

[324] *Id.*

[325] *See* Justin (Gus) Hurwitz, *Cyberensuring Security*, 49 CONN. L. REV. 1495, 1534-35 (2017) (arguing that insurance can have a significant impact on market behavior, and that strict liability can function as insurance from an economics perspective); Ram Shankar Siva Kumar & Frank Nagle, *The Case for AI Insurance*, HARV. BUS. REV. (Apr. 29, 2020), https://hbr.org/2020/04/the-case-for-ai-insurance [https://perma.cc/MM8U-CAP4]; John Buchanan, *Generative AI Loss Adds New Risk Area to Insurance Policies*, BLOOMBERG L. (May 9, 2023),

A statute could both require and create a repository scheme for these algorithms, perhaps subject to a low fee. Although the statute could include incentives for development in the future, such as limited patent exclusivity, this is not necessarily a requirement of an AI statute at this time.[326] Additional benefits could include competition: adverse parties could evaluate their competitor's AI and create better products; new entrants could create products and fulfill greater or different needs than those available.

Moreover, the AIA would have a complete list of auditable organizations, along with potential or probable issues in their respective models, crowd-sourced for efficiency. Hosting could also create centralized obligations, such as a listing of "AI indications for use," a copy of the organization's privacy policy (if applicable), and a link to report a complaint to the organization or to the AIA. Although this model would work primarily for AI developers, a similar model, consisting of a link to report issues on a licensee's website, could also ensure licensees timely report issues to AI developers. It is also crucial that individuals and organizations have fair notice both of AI's use and AI's extent of use, so that deviations from the AI's use can be reasonably avoided if desired.[327] Communicating this information could include a blend of presence in an organization's privacy notice (if AI involves personal information processing), on the AI information hosting site, and boldly on the product's actual label.[328]

---

https://news.bloomberglaw.com/us-law-week/generative-ai-loss-adds-new-risk-area-to-insurance-policies [https://perma.cc/4DQF-KXXB].

[326] *See* Tschider, *supra* note 54, at 721.

[327] Kiel Brennan-Marquez, *"Fair Notice" in the Age of AI*, *in* THE OXFORD HANDBOOK OF ETHICAL AI 754, 763 (2020).

[328] Required inclusion in labels and on the site, along with basic information disclosed about the AI model means that plaintiffs could potentially pursue misrepresentation claims or contact their state attorney general about possible fairness issues. By including information in a privacy notice and labels, the FTC could also enforce potential issues related to privacy and misleading labeling.

5. Enforcement and Private Right of Action

Much of this Article is devoted to discussion of impediments to AI recovery in litigation. While this might be justification for *ex ante* regulation, it is foolish to believe that an *ex ante* regulatory approach alone will prevent all harms resulting from poor AI design, harms that could result in mass torts.[329] Of course, no regulatory approach would be effective without enforcement. Like many regulatory approaches of its kind, the AI statute should include differentiated civil monetary penalties levied by the AIA for negligent non-compliance with the statute or for reckless or intentional behavior. Such penalties should reflect the inherently collaborative nature of AI and the potentially dangerous impacts of poor choices in design. These penalties should be substantial enough that organizations comply without being overly heavy-handed. They should be specific as to their calculation, defining the number of infractions required to be penalized and how that number is calculated (e.g., per use, per day, per routine).

An open question is whether a private right of action would be reasonable for a statute like this. On one hand, statutes like the Telephone Consumer Protection Act, which include a private right of action, have incorporated a fine structure that has dramatically changed how companies auto-dial and auto-text customers.[330] On another, statutes like HIPAA do not have a private right of action, instead leaving the Office for Civil Rights, the Department of Health and Human Services' enforcement arm, to enforce it.[331] For regulatory schemes that

---

[329] Francis A. Citera & Gretchen N. Miller, *5 Trends to Watch: 2024 Products Liability & Mass Torts*, GREENBERGTRAURIG (Jan. 4, 2024), https://www.gtlaw.com/en/insights/2024/01/published-articles/5-trends-to-watch-2024-products-liability-mass-torts [https://perma.cc/CE7U-FE3R].

[330] Shay Dvoretzky et al., *The Evolving Telephone Consumer Protection Act Landscape Post-Duguid*, SKADDEN (Apr. 2022), https://www.skadden.com/insights/publications/2022/04/quarterly-insights/the-evolving-telephone-consumer-protection-act [https://perma.cc/3ECS-LLW6].

[331] *How OCR Enforces the HIPAA Privacy & Security Rules*, U.S. DEPT. HEALTH & HUM. SERVS. (June 7, 2017), https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/examples/how-ocr-enforces-the-

permit a private right of action, some actions are preempted when a regulatory agency has engaged in lengthy *ex ante* review.[332] An important determination is what structure may be used for private rights of action. Due to the potential for substantially impactful class-action lawsuits in this area, courts may be challenged by tort reform statutes and other limitations on significant jury awards. For this reason, it may be comparatively easier to work with a statutorily established Civil Monetary Penalty model enforced by the AIA and a private right of action that permits statutory recovery but explicitly rejects nominal damages.

One case for a private right of action is for learning purposes, while another is for compensation. First, the adversarial system is highly desirable because it can illuminate certain aspects of organizational and technological behavior that may not be available to regulatory authorities or disclosed under a statutory regime. The information that comes from experts in these proceedings, as well as how evidence can characterize these practices differently is certainly valuable to understanding the bigger (and deeper) picture. Second, the adversarial process can compensate harmed people and organizations in a way that administrative agencies and private contract cannot.[333]

---

hipaa-privacy-and-security-rules/index.html [https://perma.cc/LTM5-RCT2]; Kevin D. Holden, *Fourth Circuit Court of Appeals Holds HIPAA Does Not Create a Private Right of Action*, NAT'L L. REV. (June 25, 2021), https://www.natlawreview.com/article/fourth-circuit-court-appeals-holds-hipaa-does-not-create-private-right-action [https://perma.cc/K5ZA-BQGD].

[332] *See, e.g.*, Tschider, *supra* note 111, at 1575.

[333] Although courts typically prefer contracts between organizations to serve as the basis for damages, the significant degree of information asymmetry in these contracts complicates the degree to which certain types of harms could be readily foreseeable at the time of assent. Statutory duty does not generally provide a sufficient exemption to the economic-loss doctrine. But products liability can overcome the economic-loss doctrine if harm is physical in nature. *See* Calo, *supra* note 138, at 599; *see, e.g.*, Simantob v. Mullican Flooring, L.P., 527 F. App'x 799 (10th Cir. 2013). A valid exemption to the economic loss doctrine concerns special relationships that

In the event a private right of action was available, the definition of AI as a product could help convince courts to consider causes of action in products liability. In the event courts hear these cases, whether pursuant to a private right of action or in traditional litigation, there may be ways to better align the products liability path to AI technology. Although strict liability might seem most desirable given its simplicity, it best applies to products that are continuously learning (and can be defective) when they are live or for defined high-risk products.[334] Because strict liability often applies to manufacturing defects that can only be attributed to the manufacturer and are by their nature unavoidable (and may be unforeseeable), or for ultrahazardous activities, it would logically follow that unlocked AI could introduce unavailable defects that harm users.[335] Moreover, unlocked AI's continuous learning combined with its inscrutability would make it nearly impossible to interrogate the cause of some AI function causing the harm.

In contrast, products that are locked upon completion and are not continuously learning may be stable enough to where courts could examine whether they were appropriately designed, trained, and tested, as well as how quickly AI

---

establish independent duties. *See* Robert L. Rabin, *Respecting Boundaries and the Economic Loss Rule in Tort*, 48 ARIZ. L. REV. 857, 867 (2006). These factors could potentially convince a court that the economic-loss doctrine would not be appropriate to block such a case in tort, at least in some jurisdictions. Additionally, contracts are typically between two parties, but the harm flowing from AI use will likely extend beyond contractual privity, and often other parties are foreclosed from recovery. *See* TURNER, *supra* note 152, at 110.

[334] Continuously learning AI applications will likely exhibit issues that may not be predictable or avoidable by users, creating a reasonably unsafe environment. For this reason, it may be a better choice to use strict liability for these AI. *See id.* at 95.

[335] Joseph W. Little, *The Place of Consumer Expectations in Product Strict Liability Actions for Defectively Designed Products*, 61 TENN. L. REV. 1189, 1194-97 (1994) (arguing that strict liability should be primarily available for ultrahazardous or dangerous products). In this case, although AI may or may not be "dangerous," an implementation with somewhat unlimited risk could satisfy this standard.

developers responded to issues posted on the hosting site.[336] Products that are not continuously learning likely can be litigated under the products-liability design-defect path.[337]

There are three areas where courts may need to evolve to address AI harms. First, although foreseeability usually applies to general negligence cases, it also applies in the design defect analysis. For example, the idea that the design defect even exists depends on some foreseeability as to what could go wrong and how it could affect downstream users. Because AI can introduce unusual outcomes, a design defect should be *de facto* foreseeable when the defect falls under the indications for use, as disclosed on the hosting site and likely on the AI's label (or associated documentation). This creates some limitation in foreseeability without resorting to strict liability.

Second, courts may need to offer some relaxation in the pleading standards related to reasonable alternative design. Reasonable alternative design could be one of the most difficult required plaintiff showings in products liability because much of design is kept highly confidential and is not available except in discovery.[338] For computer systems, this information will not likely be able to be reverse engineered, unlike physical products with observable parts. Information gleaned from a required hosting site, if the AI statute requires this, can provide enough information to plausibly make

---

[336] As an alternative, Bryan Choi differentiates between fault avoidance and fault tolerance as different models, with the former seeking to create high-quality code, and the latter ensuring that errors do not manifest into failures. *See* Choi, *supra* note 129, at 101-02. These differences illustrate different humans outside the loop considerations, which could be evaluated in design-defect litigation. Despite hardware redundancies, software redundancies, even for AI powering high-risk applications, are generally not developed. *Id.* at 104.

[337] *Id.* at 94. Products liability can be useful to plaintiffs when one defendant can be found. The defendant can then seek contribution from others if the fault is not 100% their own. John R. Richards & Howard S. Suskin, *When Do Indemnity Claims for Product Liability Accrue?*, 19 ANDREWS LITIG. REP., Mar. 2008, at 1, 1.

[338] Because most AI may not be protected via patent or copyright, it is likely much of the technology, including design, will fall to trade secrecy and be generally unavailable to plaintiffs. *See* Tschider, *supra* note 52, at 141.

arguments related to reasonable alternative design, though at least until discovery, courts may need to offer some flexibility as to the details of this arguments in favor of the plaintiff.[339] AI can be more understood over time and in the aggregate.[340] Plaintiffs should also be able to offer a statutory basis for products liability based on the proposed AI statute itself, allowing them to proceed if a requirement for safety, efficacy, or fairness has not been adequately met. When a statutory basis for products liability is fulfilled, such as failure to conduct risk assessments, plaintiffs should not need to demonstrate a reasonable alternative design.

Finally, one of the most significant advantages to bringing cases under a products liability theory is the existing evidentiary flexibility to demonstrate causation. Under *res ipsa loquitur*, if an AI developer creates a product that malfunctions and causes the plaintiff's harm, the plaintiff does not usually need to independently show proximate (legal) causation, so long as the harm can reasonably be attributed to the malfunction of the product.[341] This doctrine may be interpreted differently in specific jurisdictions, but the relative flexibility is hugely useful for AI harms, where it may be difficult, if not impossible, to overcome opacity issues to determine how the AI malfunctioned.

Although affirmative defenses can certainly illustrate how a plaintiff did not behave reasonably in relation to the product (which may comparatively result in an unsuccessful lawsuit), contractual limitations on liability or assumptions of risk should not be upheld based on public policy. Because an AI developer could feasibly argue that the risks were

---

[339] Not all courts treat reasonable alternative design with the same stringency. *See* Alex Purvis & Simon Bailey, *Alternative Approaches to Alternative Design*, IADC COMM. NEWS (Dec. 2014), https://www.iadclaw.org/assets/1/19/Product_Liability_Dec_2014.pdf?3059 [https://perma.cc/WKQ5-ZSTJ]. Courts may see the need in some cases to adapt a less stringent version for technologies that cannot be understood by disassembling them.

[340] *See* Powers & Ganascia, *supra* note 58, at 47.

[341] RESTATEMENT (THIRD) OF TORTS: PHYS. & EMOT. HARM § 17 (AM. L. INST. 2010). It was recommended that two factors be added to the Restatement of Torts: Products Liability. First, that the defendant has superior knowledge to the plaintiff, and second, that the defendant is most likely responsible for the accident, even if it cannot be directly ascertained.

unforeseeable when the AI is inscrutable, risk shifting to entities that have no ability anticipate or gauge risk is inherently one-sided—favoring AI developers. If AI developers can no longer rely on contractual provisions to absorb potential liability, they may instead focus on building better products, the ultimate human-centered goal.

Conclusion

This Article has examined the many reasons why a human in the loop largely benefits commercial entities while potentially harming the human. Legal conceptions of tort coupled with the use of restrictive contractual clauses, such as limitations on liability, have created a model of liability that largely insulates AI development companies. Although AI innovation is important, what is the cost?

AI is often more complex than it seems, not only because of the technical complexity of creating AI applications but also because of the wide variation in and the number of entities involved in AI's creation and use. Legally enforcing human decisions outside the loop and creating a regulatory model to standardize these obligations positions AI to be safer, fairer, and more accurate than our current liability model provides. A regulatory model establishes a foundational, common set of statutory expectations for all organizations creating AI, regardless of the organization's size. It offers the opportunity for establishing and revising requirements as the United States and organizations creating AI also learn more about these technologies.

Developing a consistent regulatory model with an AI statute that establishes baseline requirements will ensure all organizations integrate at least basic processes to protect the people who use their systems. Even if the current political landscape does not support the passage of a new AI law and a new administrative agency, courts still have an opportunity to interrogate long-standing challenges in tort litigation that will substantially frustrate plaintiff interests and case viability. Ultimately, AI products exist because humans decide to create them. Thus, the future of AI should leverage the immense

power of human intelligence to improve flourishing, but not at the expense of humans themselves.[342]

[342] *See* Powers & Ganascia, *supra* note 58, at 48. Powers and Ganascia see AI as a set of technologies that serve, are "embedded in a *system* of human agents, other artificial agents, laws, nonintelligent infrastructures, and social norms." *Id.*