

# Applying a Common Enterprise Theory of Liability to Clinical AI Systems

Benny Chan<sup>†</sup>

*The advent of artificial intelligence (“AI”) holds great potential to improve clinical diagnostics. At the same time, there are important questions of liability for harms arising from the use of this technology. Due to their complexity, opacity, and lack of foreseeability, AI systems are not easily accommodated by traditional liability frameworks. This difficulty is compounded in the health care space where various actors, namely physicians and health care organizations, are subject to distinct but interrelated legal duties regarding the use of health technology. Without a principled way to apportion responsibility among these actors, patients may find it difficult to recover for injuries. In this Article, I propose that physicians, manufacturers of clinical AI systems, and hospitals be considered a common enterprise for the purposes of liability. This proposed framework helps facilitate the apportioning of responsibility among disparate actors under a single legal theory. Such an approach responds to concerns about the responsibility gap engendered by clinical AI technology as it shifts away from individualistic notions of responsibility, embodied by negligence and products liability, toward a more distributed conception. In addition to favoring plaintiff recovery, a common enterprise strict liability approach would create strong incentives for the relevant actors to take care.*

## I. INTRODUCTION

Medicine is currently undergoing an Artificial Intelligence (“AI”) revolution.<sup>1</sup> The rise of big data and development of sophisticated machine-learning techniques holds

---

<sup>†</sup>Counsel, Department of Justice Canada (Health Legal Services Unit), Ottawa, Canada. The author would like to thank Gregory Scopino, Jennifer Anderson, Nicholson Price, and two anonymous reviewers for their enormously helpful comments. Special thanks also go to the AJLM editors for their assistance throughout the publication process. The views expressed herein should not be taken as reflecting the position of the Department of Justice. The author can be contacted at [bc742@georgetown.edu](mailto:bc742@georgetown.edu)

<sup>1</sup>Mindful of the intense and technical debate over the definition of AI, I will simply adopt Ryan Calo’s definition of AI as a “set of techniques aimed at approximating some aspect of human or animal cognition using machines” for the purposes of this Article. Ryan Calo, *Artificial Intelligence Policy: A Primer and Roadmap*, 51 U.C. DAVIS L. REV. 399, 404 (2017). For a more detailed discussion, see also Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 354, 359-362 (2016).

great potential to improve every step of the clinical process.<sup>2</sup> However, the increasing use of clinical AI systems raises the important question of how the law should deal with liability for harms arising from the use of these systems. Recently, IBM Watson recommended unsafe and incorrect cancer treatments, underscoring the danger that a flawed algorithm poses to patients.<sup>3</sup> It is imperative that tort law adapt to this technological challenge. Not only is an adequate liability regime important for victims seeking compensation, but also a flawed regime may compromise the expected benefits that AI technology holds for the health care system as a whole.<sup>4</sup>

This Article proposes a theory of liability whereby physicians, manufacturers of clinical AI systems, and hospitals that employ the systems are considered to be engaged in a common enterprise for the purposes of tort liability. As members of a common enterprise, they should furthermore be held jointly and strictly liable for harms caused by clinical AI systems. The argument put forth in this Article is an extension of David Vladeck's proposal to impose common enterprise liability ("CEL") on component manufacturers of autonomous vehicles.<sup>5</sup> By appropriating Vladeck's criteria of "common objective" to determine who is part of a common enterprise, the proposed framework can facilitate the apportioning of responsibility among disparate actors under a single legal theory. The proposed framework thereby accounts for the dispersion of responsibility among those involved in the creation, implementation, and operation of clinical AI systems.

This proposal draws inspiration from the scholarly literature on hospital enterprise liability ("EL"). In particular, it mobilizes the insight that the hospital acts as a locus of accountability for patient safety to justify the hospital's inclusion in the common enterprise. Medical errors are often the result of faulty systems implemented by health care organizations.<sup>6</sup> The introduction of new technology such as AI increases the risk of new errors, and it falls on health care organizations to implement proper procedures and surveillance systems to prevent such risks from materializing. This has long been implicit in the doctrine of hospital corporate liability, which holds that the hospital has the duty to use reasonable care in the maintenance of safe and adequate equipment.<sup>7</sup> Holding hospitals liable will also incentivize greater attention being paid to patient safety in the clinical use of AI technology.

This Article proceeds in three parts. Part II consists of a brief overview of AI's potential to revolutionize medicine. In the near future, AI systems may acquire the capacity

---

<sup>2</sup>See, e.g., Joachim Roski et al., *How Artificial Intelligence is Changing Health and Health Care*, in *ARTIFICIAL INTELLIGENCE IN HEALTH CARE: THE HOPE, THE HYPE, THE PROMISE, THE PERIL* 64 (Michael Matheny et al. ed., 2019) [hereinafter *ARTIFICIAL INTELLIGENCE IN HEALTH CARE*].

<sup>3</sup>Casey Ross & Ike Swetlitz, *IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close*, STAT (Sept. 5, 2017), <https://www.statnews.com/2017/09/05/watson-ibm-cancer/> [<https://perma.cc/P6CP-HUY6>].

<sup>4</sup>See EUR. COMM'N, EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES, *LIABILITY FOR ARTIFICIAL INTELLIGENCE AND OTHER EMERGING DIGITAL TECHNOLOGIES* 11 (2019); RYAN ABBOTT, *THE REASONABLE ROBOT* 53 (2020) (noting that tort law "has far-reaching and sometimes complex impact on behavior," including on the introduction and use of new technologies).

<sup>5</sup>David Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 129 n.39 (2014).

<sup>6</sup>See INST. MED., *TO ERR IS HUMAN: BUILDING A SAFER HEALTH SYSTEM* 58-59 (2000) (noting that complex and tightly coupled systems are more prone to accidents. Complex systems have many specialized and interdependent parts such that if one part that serves multiple functions fails, all of the dependent functions fail as well. In systems that are tightly coupled, processes are more time-dependent and sequences are more fixed. As such, things can unravel quickly which makes it difficult to intercept errors and prevents speedy recovery from events. Activities in emergency rooms, surgical suites, and intensive care units are examples of complex and tightly coupled systems).

<sup>7</sup>See *Darling v. Charleston Cmty. Mem'l Hosp.*, 211 N.E.2d 253, 257 (Ill. 1965).

to perform a clinical diagnosis without the intervention of human clinicians.<sup>8</sup> Underlying AI's potential is the application of powerful machine learning algorithms to generate accurate predictions.<sup>9</sup> Despite AI's great promise, the technology faces significant limitations that can lead to inaccurate results.<sup>10</sup> Furthermore, the technology's lack of ability to explain its diagnoses or recommendations will make it difficult for physicians to evaluate an AI system's output against their own expertise.

Part III undertakes a critical analysis of multiple liability frameworks that have been proposed for AI systems: products liability, negligence, agency law, AI personhood, EL, and CEL. The first four frameworks have received the most scholarly attention in discussions of AI liability, though all of these encounter significant problems and uncertainties in their application to clinical AI systems. While products liability has been considered a natural liability framework for harms arising from AI systems, there are numerous doctrinal obstacles that stand in the way.<sup>11</sup> Moreover, the difficulty of proving a defect and the application of the learned intermediary ("LI") principle add uncertainty regarding a plaintiff's ability to recover from the manufacturer.<sup>12</sup> Negligence suffers from a similar weakness by requiring the plaintiff to prove causation and fault, which may prove burdensome given AI's explainability problem. Both products liability and negligence are also limited in their ability to apportion liability among disparate actors. While agency law can account for multiple tortfeasors (as co-principals) and accommodate the lack of foreseeability of AI actions, the control requirement between principal and agent makes it too limited of a legal theory to account for the diverse ways in which actors can contribute to AI harm. Proposals to recognize AI as persons remain contentious due to, among other reasons, the ethical implications.<sup>13</sup>

Part IV consists of arguments for why a CEL-based approach to clinical AI systems is both reasonable and desirable. One can plausibly consider physicians, manufacturers, and hospitals to be engaged in a common enterprise as clinical AI systems are designed for use by health care professionals and organizations to provide care to patients. Therefore, there is a strong conceptual overlap in the objectives pursued by each actor. As for the advantages of a CEL-based approach, a single theory of liability for AI harms resonates with concerns about the responsibility gap engendered by clinical AI technology. Moreover, the imposition of strict liability is responsive to the realities of modern tort litigation and generates a fairer outcome given the benefits that physicians, manufacturers, and health care organizations derive from the use of clinical AI systems. This Article ends with a discussion about future legal reforms aimed at implementing a CEL-based state-level approach, with the possibility of incorporating limited AI personhood into the proposed framework.

---

<sup>8</sup>See Timothy Craig Allen, *Regulating Artificial Intelligence for a Successful Pathology Future*, 143 ARCHIVES PATHOLOGY & LAB'Y MED. 1175, 1177 (2019).

<sup>9</sup>See, e.g., Fei Jiang et al., *Artificial intelligence in Healthcare: Past, Present and Future*, 2 STROKE & VASCULAR NEUROLOGY 230, 230 (2017).

<sup>10</sup>See, e.g., Kee Yuan Ngiam & Ing Wei Khor, *Big Data and Machine Learning Algorithms for Health-Care Delivery*, 20 LANCET ONCOLOGY e262, e266 (2019).

<sup>11</sup>See, e.g., ABBOTT, *supra* note 4, at 132 (noting that products liability law may require that AI be a commercial product and not a service); Nicolas Terry, *Of Regulating Healthcare AI and Robots*, 18 YALE J. HEALTH POL'Y, L., & ETHICS 133, 162 (2019).

<sup>12</sup>See *infra* notes 72-79 and accompanying text.

<sup>13</sup>See e.g., Joanna J. Bryson, Mihailis E. Diamantis & Thomas D. Grant, *Of, for, and by the People: The Legal Lacuna of Synthetic Persons*, 25 A.I. & L. 273, 275 (2017) (opposing the extension of rights to algorithms partly due to the implications on human rights); citations *infra*, note 150.

## II. BACKGROUND: THE PROMISE AND PERIL OF CLINICAL AI

The development of AI technology has led to a paradigm shift in medicine. In the long run, AI has the potential to push the boundaries of what human health providers can do and provide a tool to manage patients and medical resources.<sup>14</sup> Importantly, AI technology is predicted to have a significant impact on each step of the clinical process.<sup>15</sup> AI has the potential to improve medical prognosis by using thousands of predictor variables taken from electronic records and other sources, unlike human providers who rely on scoring tools.<sup>16</sup> Moreover, clinical decision-support systems based on AI algorithms will soon be able to “tailor relevant advice to specific decisions and treatment recommendations.”<sup>17</sup>

Timothy Craig Allen predicts that the future of clinical AI systems will develop in three phases.<sup>18</sup> In the first (“present”) phase, the physician is “in the loop” in being fully in control of the AI system. Here, the AI system serves merely as “another tool for diagnosis” that renders the clinician more efficient. In the second (“near future”) phase, the clinician moves from being “in the loop” to being “on the loop,” as the AI system now has the capacity to perform a clinical diagnosis and issue a report without the clinician’s review—though the clinician may still want to exercise quality control over the results. The AI system in this phase will have drastically reduced—if not entirely eliminated—human involvement. In the third (“distant future”) phase, the AI system becomes autonomous, and the clinician is taken “out of the loop” entirely. Here, a human clinician is “entirely unnecessary to render an actionable diagnosis or to institute treatment.”<sup>19</sup> The second and third phases, if they come to fruition, would require a radical rethinking of existing systems of liability and regulation.

Underlying AI’s powerful potential in the field of health care is the use of sophisticated machine learning algorithms to extract insights “from a large volume of health care data, and then use the obtained insights to assist clinical practice.”<sup>20</sup> Clinical machine learning tools are generally based on supervised machine learning (“ML”),<sup>21</sup> which refers to techniques in which a model is trained on datasets consisting of a range of inputs (or features) that are associated with a known outcome.<sup>22</sup> When applied to new data, a well-trained algorithm will uncover patterns or structures that are implicitly present, thereby allowing for accurate predictions to be made. Supervised ML algorithms are iteratively retrained to improve their predictive accuracy using an optimization

<sup>14</sup>See, e.g., W. Nicholson Price II, *Artificial Intelligence in the Medical System: Four Roles for Potential Transformation*, 18 *YALE J. HEALTH POL’Y, L., & ETHICS* 124, 124-25 (2019); ABBOTT, *supra* note 4, at 56 (“Why should AI not be able to outperform a person when the AI can access the entire wealth of medical literature with perfect recall, benefit from the experience of directly having treated millions of patients, and be immune to fatigue?”).

<sup>15</sup>ARTIFICIAL INTELLIGENCE IN HEALTH CARE, *supra* note 2, at 64.

<sup>16</sup>CANADIAN MED. ASS’N, *THE FUTURE OF TECHNOLOGY IN HEALTH AND HEALTH CARE: A PRIMER* 10 (2018), <https://www.cma.ca/sites/default/files/pdf/health-advocacy/activity/2018-08-15-future-technology-health-care-e.pdf> [<https://perma.cc/UU2A-32U2>].

<sup>17</sup>Stephan Fihn et al., *Deploying AI in Clinical Settings*, in *ARTIFICIAL INTELLIGENCE IN HEALTH CARE*, *supra* note 2, at 154.

<sup>18</sup>Allen, *supra* note 8, at 1177. While Allen is writing specifically for pathologists, his insights are readily generalizable to other areas of medicine.

<sup>19</sup>*Id.*

<sup>20</sup>Jiang et al., *supra* note 9, at 230.

<sup>21</sup>Ngiam & Khor, *supra* note 10, at e266.

<sup>22</sup>For tasks such as image recognition or language processing, a feature selector will need to first process the variables by “picking out identifiable characteristics from the dataset which then can be represented in a numerical matrix and understood by the algorithm.” Jenni A. M. Sidney-Gibbons & Chris J. Sidney-Gibbons, *Machine Learning in Medicine: A Practical Introduction*, 19 *BMC MED. RSCH. METHODOLOGY* 1, 2 (2019).

technique.<sup>23</sup> A special kind of ML algorithm known as deep neural networks has been gaining in popularity due to its ability to accurately label objects in images.<sup>24</sup> Neural networks hold great promise in their potential application to image-based medical sub-fields such as radiology, ophthalmology, dermatology, and pathology.<sup>25</sup>

Despite its great promise, ML algorithms currently face significant limitations which can result in inaccurate clinical predictions or recommendations.<sup>26</sup> Among these limitations is the potential for ML algorithms to accidentally exploit an unknown and possibly unreliable confounding variable. This could negatively affect the algorithm's applicability to new data sets. For example, unreliable confounders exploited by deep learning models have contributed to inaccurate detections of melanoma and hip fractures.<sup>27</sup> There are also challenges relating to the generalizability of AI findings to new populations due to factors such as technical differences between sites and variations in administrative practices.<sup>28</sup> Generalizability has been called the "Achilles heel" of deep learning, given that "algorithms trained on a specific data set often perform very well on similar data, but often may yield poor performance in the case of data that have not been seen in the training process."<sup>29</sup> Studies have shown that even subtle differences between populations can greatly affect a clinical AI system's predictive accuracy.<sup>30</sup>

Recent reports on the underperformance of IBM's cancer AI algorithm, called Watson for Oncology ("WFO"), provides a good example of the limitations and pitfalls that future clinical AI systems are likely to encounter. In the years following WFO's well-publicized victory over two human contestants on the gameshow *Jeopardy!*, IBM advertised and sold WFO to doctors around the world as a platform with the capacity to recommend the best cancer treatments for specific patients.<sup>31</sup> However, a 2018 STAT report revealed statements from company specialists and customers that WFO had generated "multiple examples of unsafe and incorrect treatment recommendations."<sup>32</sup> For example, WFO had recommended bevacizumab (Avastin) to a patient with evidence of severe bleeding, despite a clear contraindication and a warning from the FDA. There were purported flaws in the methods used to train WFO, including the small number of cases used as inputs.<sup>33</sup> While fortunately no patients were harmed by WFO's underperformance, this case underlines the real danger that a flawed algorithm can pose to patients.

---

<sup>23</sup>See *id.* at 3.

<sup>24</sup>These algorithms "consist of layers of nodes that each use simple mathematical operations to perform a specific operation on the activation of the layer before, leading to the emergence of increasingly abstract representations of the input image." Thomas Grote & Philipp Berens, *On the Ethics of Algorithmic Decision-Making in Healthcare*, 46 J. MED. ETHICS 205, 206 (2020).

<sup>25</sup>See Thomas Davenport & Ravi Kalakota, *The Potential for Artificial Intelligence in Healthcare*, 6 FUTURE HEALTHCARE J. 94, 94 (2019). Indeed, it is predicted that AI's capacity to interpret digitized images (e.g. x-rays) will eventually outstrip that of human radiologists and pathologists. See, e.g., CANADIAN MED. ASS'N, *supra* note 16, at 10.

<sup>26</sup>See, e.g., Ngiam & Khor, *supra* note 10, at e266.

<sup>27</sup>See, e.g., Christopher J. Kelly et al., *Key Challenges for Delivering Clinical Impact with Artificial Intelligence*, 17 BMC MED. 1, 4 (2019).

<sup>28</sup>*Id.*

<sup>29</sup>Michael P. Recht et al., *Integrating Artificial Intelligence into the Clinical Practice of Radiology: Challenges and Recommendations*, 30 EUR. RADIOLOGY 3576, 3579 (2020).

<sup>30</sup>Fihn et al., *supra* note 17, at 166.

<sup>31</sup>Ross & Swetlitz, *supra* note 3.

<sup>32</sup>Casey Ross & Ike Swetlitz, *IBM's Watson Supercomputer Recommended 'Unsafe and Incorrect' Cancer Treatments, Internal Documents Show*, STAT (July 25, 2018), <https://www.statnews.com/wp-content/uploads/2018/09/IBMs-Watson-recommended-unsafe-and-incorrect-cancer-treatments-STAT.pdf> [<https://perma.cc/WD77-YBGA>].

<sup>33</sup>*Id.*

AI algorithms' lack of explainability will also be an obstacle for the clinical application of AI technology. The term explainability (often used interchangeably with "interpretability") refers to that "characteristic of an AI-driven system allowing a person to reconstruct why a certain AI came up with the presented predictions."<sup>34</sup> Current ML algorithms do not provide an explanation or justification of why a certain result was generated due to the opacity of ML algorithms' complex form of mathematical representation.<sup>35</sup> As such, human users are confronted with what has been called the "black box" problem—defined as an "inability to fully understand an AI's decision-making process and the inability to predict the AI's decisions or outputs."<sup>36</sup> Computer scientist Geoffrey Hinton, a pioneer in deep learning, explains why there is no simple explanation for how a neural net arrives at a specific result:

Understandably, clinicians, scientists, patients, and regulators would all prefer to have a simple explanation for how a neural net arrives at its classification of a particular case. In the example of predicting whether a patient has a disease, they would like to know what hidden factors the network is using. However, when a deep neural network is trained to make predictions on a big data set, it typically uses its layers of learned, nonlinear features to model a huge number of complicated but weak regularities in the data. It is generally infeasible to interpret these features because their meaning depends on complex interactions with uninterpreted features in other layers.<sup>37</sup>

One consequence of this lack of explainability is that it becomes difficult for clinicians and health care organizations to evaluate product quality in the marketplace.<sup>38</sup> Unlike drugs and other medical technologies, algorithms are not normally conducive to verification through clinical trials.<sup>39</sup> Black-box models also renders it difficult to identify

---

<sup>34</sup>Julia Amann et al., *Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective*, 20 BMC MED. INFORMATICS & DECISION MAKING 1, 2 (2020). The authors note that explainability has many facets and is not clearly defined. Other authors distinguish explainability from interpretability. See, e.g., Boris Babic et al., *Beware Explanations from AI in Health Care*, 373 SCIENCE 284, 284 (2021).

<sup>35</sup>Grote & Berens, *supra* note 24, at 207.

<sup>36</sup>Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J. L. & TECH. 889, 905 (2018). Bathaee further divides the black box problem into the categories of Strong Black Boxes and Weak Black boxes. The former refers to AI decision-making processes that are entirely opaque to human beings in that there is no way to "determine (a) how the AI arrived at a decision or prediction, (b) what information is outcome determinative to the AI, or (c) to obtain a ranking of the variables processed by the AI in the order of their importance." *Id.* at 906. The latter refers to AI decision-making processes that are opaque but susceptible to reverse engineering or probing to "determine a loose ranking of the importance of the variables the AI takes into account." *Id.*

<sup>37</sup>Geoffrey Hinton, *Deep Learning – A Technology with the Potential to Transform Health Care*, 320 JAMA 1101, 1102 (2018).

<sup>38</sup>See W. Nicholson Price II, *Regulating Black-Box Medicine*, 116 MICH. L. REV. 421, 443 (2017) (noting that machine-learning methods are difficult to evaluate due to algorithmic opacity and complexity).

<sup>39</sup>See Roger A. Ford & W. Nicholson Price II, *Privacy and Accountability in Black-Box Medicine*, 23 MICH. TELECOM. & TECH. L. REV. 1, 16 (2016) (noting that black-box algorithms identify interventions that are specific to an individual, not a cohort of patients with similar indications. The inability to randomize across a sample population and observe different outcomes makes it impossible to predict individual responses of individual patients through a clinical trial); Thomas M. Maddox et al., *Questions for Artificial Intelligence in Health Care*, 321 JAMA 31 (2019) (suggesting that the "use of deep learning and other analytic approaches in AI adds an additional challenge. Because these techniques, by definition, generate insights via unobservable methods, clinicians cannot apply the face validity available in more traditional clinical decision tools"); Derek C. Angus, *Randomized Clinical Trials of Artificial Intelligence*, 323 JAMA 1043 (2020) (commenting on the complications and uncertainties involved in conducting randomized clinical trials on AI-enabled decision support tools that are continually learning).

and correct biases such as errors among patients belonging to underrepresented or marginalized groups.<sup>40</sup> Bias is difficult to avoid due to imperfect training data; even a theoretically fair model can in practice be biased upon interacting with the larger healthcare system.<sup>41</sup>

Moreover, it will be difficult for physicians to evaluate the soundness of an AI system's diagnosis or recommendation against their own knowledge; both will be considered experts of sorts, yet with different training and distinct ways of reasoning.<sup>42</sup> Some argue that this inability of physicians to vet the quality of training labels and data is contrary to evidence-based medicine.<sup>43</sup> The use of black box medicine also raises questions of patient autonomy and informed consent as patients are unable to question the AI system. Without the benefit of a justification for an AI-generated recommendation that is understandable to a human user, patients may be deprived of the opportunity to engage in a dialogue with their clinicians about the underlying reasoning. This in turn places patients in a situation where they have to make important health care decisions without sufficient information.<sup>44</sup> For these and other reasons, there have been strong calls for explainability in clinical AI systems.<sup>45</sup> However, greater AI explainability may come with trade-offs in the form of having to limit an AI system's complexity and, in doing so, its performance.<sup>46</sup> There are also questions regarding the inherent limitations of AI explainability techniques.<sup>47</sup>

### III. CURRENT TORT FRAMEWORKS: A CRITICAL ANALYSIS

While there has yet to be any case law involving the use of clinical AI systems or algorithms, this is likely to change with the increasing use of clinical AI systems.<sup>48</sup>

---

<sup>40</sup>Thomas Quinn et al., *The Three Ghosts of Medical AI: Can the Black-Box Present Deliver?*, A.I. MED. (forthcoming 2021) (manuscript at 3).

<sup>41</sup>*Id.* For instance, an adaptive algorithm can learn from racial, ethnic, and socioeconomic disparities in care and outcomes that pervade a healthcare system. The predictions or recommendations generated by the algorithm then reinforce these biases, creating a negative feedback loop. Clinicians may inadvertently contribute to this feedback loop if, owing to time pressure or fear of liability, they treat the algorithm's recommendations as infallible and thereby fail to notice or correct for the biased outputs. Matthew DeCamp & Charlotta Lindvall, *Latent Bias and the Implementation of Artificial Intelligence in Medicine*, 27 J. AM. MED. INFORMATICS. ASS'N 2020, 2021 (2020).

<sup>42</sup>Grote & Berens, *supra* note 24, at 207.

<sup>43</sup>*See, e.g.*, Shinjini Kundu, *AI in Medicine Must be Explainable*, 27 NATURE MED. 1328, 1328 (2021).

<sup>44</sup>*Id.*, *see also* Amann et al., *supra* note 34, at 6 (arguing that “[s]ince clinicians are no longer able to fully comprehend the inner workings and calculations of the decision aid they are not able to explain to the patient how certain outcomes or recommendations are derived”).

<sup>45</sup>*See, e.g.*, Kelly et al., *supra* note 27, at 5 (holding that explainability would improve “experts’ ability to recognize system errors, detect rules based upon inappropriate reasoning, and identify the work required to remove bias”). Outside of the clinical context, there are important purposes served by AI explainability such as helping affected parties understand why a decision was made, providing grounds to contest adverse decisions, understanding how to achieve a desired result in the future. *See* Sandra Wachter, Brent Mittelstadt, & Chris Russell, *Counterfactual Explanations without Opening the Blackbox: Automated Decisions and the GDPR*, 31 HARV. J.L. & TECH. 841, 843 (2018).

<sup>46</sup>*See, e.g.*, Daniel Schönberger, *Artificial Intelligence in Healthcare: A Critical Analysis of The Legal and Ethical Implications*, 27 INT’L J.L. & INFO. TECH. 171, 195 (2019); Babic et al., *supra* note 34, 285-286 (arguing that explainable AI will produce few benefits and incur additional costs such as being misleading in the hands of imperfect users and underperforming in some tasks); ABBOTT, *supra* note 4, at 33 (“Even if theoretically possible to explain an AI outcome, it may be impracticable given the complexity of AI, the possible resource-intensive nature of such inquiries, and the need to maintain earlier versions of AI and specific data”).

<sup>47</sup>*See, e.g.*, Jeremy Petch et al., *Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology*, CANADIAN J. CARDIOLOGY (forthcoming) (arguing that “the nature of explanations as approximations may omit important information about how black box models work and why they make certain predictions”).

<sup>48</sup>W. Nicholson Price II, Sara Gerke & I. Glenn Cohen, *Potential Liability for Physicians Using Artificial Intelligence*, 322 JAMA 1765, 1765 (2019).

A coherent, fair, and predictable liability regime is important to patients, health professionals, and health organizations.<sup>49</sup> As noted in the European Commission's Expert Group on Liability and New Technologies' ("European Commission") report *Liability for Artificial Intelligence*, inadequacies in a system of liability might "compromise the expected benefits" of such a technology.<sup>50</sup> Tort liability also encourages physicians/hospitals to safely use AI and manufacturers to be diligent in their design. The following Section will critically examine some of the major liability frameworks that have been proposed for harms arising out of the operation of AI systems.

#### A. PRODUCTS LIABILITY

Products liability has received attention as a natural framework for AI liability.<sup>51</sup> The Restatement (Third) of Products Liability holds that "[o]ne engaged in the business of selling or otherwise distributing products who sells or distributes a defective product is subject to liability for harm to persons or property caused by the defect."<sup>52</sup> This seems to apply rather straightforwardly to AI systems, which are, after all, "products manufactured, distributed, and sold to customers."<sup>53</sup> Products liability requires the existence of a defect. Under a products liability framework, plaintiffs can recover for injuries resulting from defective design, defective manufacture, or defective warning from the manufacturer.<sup>54</sup> For a patient harmed by an AI system, a products liability framework presents certain advantages. Not only do manufacturers have deeper pockets than physicians, but also products liability law has a history of pro-plaintiff bias.<sup>55</sup> Accordingly, this framework has been proposed for AI systems such as autonomous vehicles.<sup>56</sup>

Nevertheless, the applicability of products liability remains a contentious issue among scholars. For one, it is unclear whether an AI algorithm would fall within the scope of products liability. The law has traditionally held that only personal property in tangible form can be considered "products."<sup>57</sup> Some commentators have argued that this is not necessarily an obstacle to the application of AI algorithms embedded in hardware because there is some legal basis for treating information integrated with a physical object as a

<sup>49</sup>See, e.g., Allen, *supra* note 8, at 1177.

<sup>50</sup>EUR. COMM'N, *supra* note 4, at 11.

<sup>51</sup>Matthew Scherer notes from "anecdotal discussions with other lawyers, the most commonly held view is that the traditional rules of products liability will apply to A.I. systems that cause harm." Matthew U. Scherer, *Of Wild Beasts and Digital Analogues: The Legal Status of Autonomous Systems*, 19 NEV. L. J. 259, 280 (2019); see also Xavier Frank, Note, *Is Watson for Oncology Per Se Unreasonably Dangerous?: Making a Case for How to Prove Products Liability Based on a Flawed Artificial Intelligence Design*, 45 AM. J.L. & MED. 273, 281, 284 (2019) (arguing that products liability would be the only viable option for a plaintiff injured by Watson Oncology to bring a suit against IBM).

<sup>52</sup>See RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 1 (AM. L. INST. 1998); cf. RESTATEMENT (SECOND) OF TORTS § 402A(2)(a) (AM. L. INST. 1965) (holding manufacturer liable for unreasonably dangerous defects in their products regardless of whether the manufacturer "exercised all possible care in the preparation and sale of the product").

<sup>53</sup>Omri Rachum-Twaig, *Whose Robot Is It Anyway?: Liability for Artificial-Intelligence Based Robots*, 2020 U. ILL. L. REV. 1141, 1154 (2020).

<sup>54</sup>See RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 2 (AM. L. INST. 1998).

<sup>55</sup>See, e.g., Willis L. M. Reese, *Products Liability and Choice of Law: The United States Proposals to the Hague Conference*, 25 VAND. L. REV. 29, 35 (1972) (noting that "the trend in the law of products liability in nearly all nations of the world has been to favor the plaintiff by imposing increasingly strict standards of liability upon the supplier").

<sup>56</sup>See, e.g., John Villasenor, *Products Liability Law as a Way to Address AI Harms*, BROOKINGS (Oct. 31, 2019), <https://www.brookings.edu/research/products-liability-law-as-a-way-to-address-ai-harms/> [<https://perma.cc/2NM6-ZZW9>].

<sup>57</sup>See RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 19(a) (AM. L. INST. 1998).



product.<sup>58</sup> Others, in contrast, see this as a serious doctrinal impediment when it comes to applying products liability law to AI systems.<sup>59</sup> Added to this complexity is the fact that software has traditionally been considered by the law to be a service and thereby falls outside the reach of products liability law.<sup>60</sup> However, this distinction may ultimately prove to be untenable given the permanent interaction between products and services in the case of AI systems.<sup>61</sup>

Even if products liability law could be clarified or broadened to include AI systems, there may be compelling policy reasons not to go down this path. Under a products liability framework, the plaintiff bears the burden of proving a defect, which may be difficult to establish when it comes to harms arising from the operation of AI systems.<sup>62</sup> Samir Chopra and Laurence White note that

[i]n the absence of a clear manufacturing defect, this will involve trying to persuade the court of a design defect or a failure to warn the user of dangers inherent in a product. In states that have adopted the Restatement (Third) of Products Liability, proving a design defect requires proof of a reasonable design alternative - a difficult challenge in a highly technical field.<sup>63</sup>

And while the Third Restatement allows for an inference of a defect, the plaintiff still bears the considerable burden of proving that the harm suffered was of a kind that would ordinarily occur as a result of a product defect and not solely the result of other causes at the time of the product's sale or distribution.<sup>64</sup> If the injury results from the AI system's machine learning capabilities, this will greatly complicate the analysis.<sup>65</sup>

There is also the practical difficulty of meeting products liability's causation requirement.<sup>66</sup> This was evident in the case of *Mracek v Bryn Mawr Hospital*,<sup>67</sup> a case

<sup>58</sup>See e.g. Karni A. Chagal-Feferkorn, *Am I an Algorithm or a Product?: When Products Liability Should Apply to Algorithmic Decision-Makers*, 30 STAN. L. & POL'Y REV. 61, 83-84 (2019) (noting that courts "treat[ed] information as a product and applied products liability laws when errors in the information caused damage, especially when the information was integrated with a physical object").

<sup>59</sup>See, e.g., Terry, *supra* note 11 at 162.

<sup>60</sup>See, e.g., Scherer, *supra* note 1, at 390 (describing this as a "thorny" issue); Iria Giuffrida, *Liability for AI Decision-Making: Some Legal and Ethical Considerations*, 88 FORDHAM L. REV. 439, 444-45, 445 n.34 (2019). *But see* Vladeck, *supra* note 5, at 132-33 n. 52 (noting products liability cases that involve allegations of software defects).

<sup>61</sup>See EUR. COMM'N, *supra* note 4, at 28.

<sup>62</sup>See, e.g., Vladeck, *supra* note 5, at 135-36; ABBOTT, *supra* note 14, at 132.

<sup>63</sup>SAMIR CHOPRA & LAURENCE F. WHITE, *A LEGAL THEORY FOR AUTONOMOUS ARTIFICIAL AGENTS* 144 (Univ. of Mich. Press ed., 2011).

<sup>64</sup>RESTATEMENT (THIRD) OF TORTS: PROD. LIAB. § 3 (AM. L. INST. 1998) ("It may be inferred that the harm sustained by the plaintiff was caused by a product defect existing at the time of sale or distribution, without proof of a specific defect when the incident that harmed the plaintiff: (a) was of a kind that ordinarily occurs as a result of product defect; and (b) was not, in the particular case, solely the result of causes other than product defect existing at the time of sale or distribution"); *see also* Vladeck, *supra* note 5, at 128 n.36 (noting that "[a]n otherwise inexplicable failure, which is not fairly described as 'ordinary,' would likely not qualify under this standard.").

<sup>65</sup>See Woodrow Barfield, *Liability for Autonomous and Artificially Intelligent Robots*, 9 PALADYN, J. BEHAV. ROBOTICS 193, 196 (2018) (noting that "the law as currently established may be useful for determining liability for mechanical defects, but not for errors resulting from the autonomous robot's 'thinking'; this is a major flaw in the current legal approach to autonomous robots").

<sup>66</sup>*Id.* ("Additionally, with intelligent and autonomous robots controlled by algorithms, there may be no design or manufacturing flaw that served as a causative factor in an accident, instead the robot involved in an accident could have been properly designed, but based on the structure of the computing architecture, or the learning taking place in deep neural networks, an unexpected error or reasoning flaw could have occurred").

<sup>67</sup>610 F. Supp. 2d 401 (E.D. Penn. 2009), *aff'd*, 363 Fed. Appx. 925, 927 (3d Cir. 2010).

involving injuries caused by the da Vinci surgical robot.<sup>68</sup> Here, the allegation was that the robot malfunctioned during a prostatectomy procedure when it flashed ‘error’ messages. This malfunction allegedly resulted in the plaintiff suffering severe injuries. The Third Circuit upheld the summary judgment against the plaintiff for a failure to adduce expert evidence permitting him to prove that the defect caused the injury.<sup>69</sup> The plaintiff faced an uphill battle from the beginning as the pool of independent experts with knowledge of this novel and proprietary technology was very limited.<sup>70</sup> While *Mracek* was a robotics case, one can imagine a similar fact pattern in the case of a novel AI diagnostic system where the only experts are employees or former employees of the manufacturer.

Another potential obstacle to the application of product liability law to AI is the LI doctrine. This doctrine holds that manufacturers of drugs and medical devices do not have a legal obligation to warn a patient about risks associated with their product when the manufacturer has already provided a warning to the doctor.<sup>71</sup> It has been applied by courts as “a blanket exemption from a duty to warn the consumer of a prescription drug of the potential dangers of a drug.”<sup>72</sup> The rationale behind the LI doctrine is that manufacturers should be able to rely upon physicians to pass warnings regarding medical products to their patients.<sup>73</sup> The extent to which the LI doctrine would apply to clinical AI systems remains an unsettled question. On one hand, some commentators have posited that the LI doctrine could shield AI manufacturers from liability for harm arising out of the use of their products.<sup>74</sup> On the other hand, courts have also recognized that if the physician does not play an active role with regard to the product and patient, then the manufacturer is precluded from invoking the LI doctrine as a defense against liability.<sup>75</sup> This suggests that whether the physician is considered a learned intermediary under product liability

---

<sup>68</sup>The da Vinci system consists of a control console unit and four slave manipulators, three for telemanipulation of surgical instruments and one for the endoscopic camera. Functionally speaking, the system allows a surgeon to visualize the surgical field using the endoscope connected to a 3D display and transforms the surgeon’s hand movement to that of the surgical instruments. C. Freschi et al., *Technical Review of the da Vinci Surgical Telesystem*, 9 INT. J. MED. ROBOTICS & COMPUTER ASSISTED SURGERY 396, 397 (2012).

<sup>69</sup>363 Fed. Appx. 925 at 926-927.

<sup>70</sup>The trial court did not allow a physician who had experience with robotic surgery to testify on the basis of insufficient, technical knowledge of the da Vinci system. See Margo Goldberg, Note, *The Robotic Army Went Crazy! The Problem of Establishing Liability in a Monopolized Field*, 38 RUTGERS COMPUTER & TECH. L. J. 225, 248 (2012).

<sup>71</sup>DAN B. DOBBS, THE LAW OF TORTS § 466 (2001).

<sup>72</sup>Timothy S. Hall, *Reimagining the Learned Intermediary Rule for the New Pharmaceutical Marketplace*, 35 SETON HALL L. REV. 193, 217 (2004).

<sup>73</sup>*Id.* at 216-17.

<sup>74</sup>See, e.g., Jessica S. Allain, *From Jeopardy! to Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems*, 73 LA. L. REV. 1049, 1069 (2013) (positing that the doctrine would “eliminat[e] any duty the manufacturer may have had directly to the patient”); W. Nicholson Price II, *Artificial Intelligence in Health Care: Applications and Legal Issues*, 14 SCITECH LAW. 10, 13 (2017) (querying whether the LI doctrine should “bow to the recognition that doctors cannot fully understand all the technologies they use or the choices such technologies help them make when they are not provided the needed and/or necessary information”).

<sup>75</sup>See *MacDonald v Ortho Pharm. Corp.*, 475 N.E.2d 65, 69, 71 (Mass. 1985). In this case, the Supreme Court of Massachusetts ruled that the physician who prescribed oral contraception to a patient was “relegated to a relatively passive role” and therefore the drug manufacturer could not invoke the LI doctrine as a defense. *Id.* at 69. Specifically, the pharmaceutical company failed to mention the risk of strokes associated with use of its contraceptive pill in the booklet distributed to patients per FDA requirements. The underlying idea here seems to be that the law will refuse to recognize the physician as an intermediary between the manufacturer and a patient in situations where the physician is not bringing her expertise and discretion to bear in consultations with the patient.

law may ultimately depend on the level of interaction between the physician and the AI system.<sup>76</sup>

A further complication is that the LI doctrine is concerned with the duty to warn. AI's unpredictability means that many AI-related risks are not foreseeable by the manufacturer, irrespective of the level of physician interaction. This differs from a case where the risk of a particular drug (e.g., oral contraception) is known to the company through the results of clinical trials. Moreover, the doctrine of the duty to warn is "premised on an unequal access to information i.e., that the manufacturer (for example) knows more about the risks than a relatively unsuspecting consumer."<sup>77</sup> However, this theory breaks down where the product acts autonomously and unpredictably. A court might find it unfair to hold an AI manufacturer liable for failure to warn about risks that were not and could not have been known. As such, uncertainty remains as to whether the LI doctrine can be invoked to shield an AI manufacturer from liability.

## B. NEGLIGENCE

Negligence frequently features as a candidate framework for addressing AI harms, as it is the default for cases of medical malpractice.<sup>78</sup> For a claim in negligence to succeed, the plaintiff must establish four elements: (1) the existence of a legal duty; (2) the breach of that duty by the defendant; (3) an injury suffered by the plaintiff; and (4) causation.<sup>79</sup> Under traditional malpractice law, a physician or other health professional could be held liable in negligence for harmful medical errors that fall below the standard of care ("the malpractice standard").<sup>80</sup> While the standard of care has traditionally been determined by reference to customary medical practice,<sup>81</sup> state courts have increasingly turned to the reasonable physician standard: what a physician with the same kind of technical background, training, and expertise as the defendant would have done in a similar situation.<sup>82</sup> The reasonable physician standard is, at least in principle, less deferential to physicians as it affords courts greater latitude in reviewing medical knowledge and custom in determining the applicable standard of care. In practice, the differences between the standards are subtle with a tendency for the two to overlap as the question of what a reasonable physician would do is often determined by reference to the customary practice of a local or national comparison group of physicians.<sup>83</sup>

---

<sup>76</sup>See Zach Harned, Matthew P. Lungren & Pranav Rajpurkar, Comment, *Machine Vision, Medical AI, and Malpractice*, HARV. J.L. & TECH. DIG. 1, 9 (2019), <https://jolt.law.harvard.edu/digest/machine-vision-medical-ai-and-malpractice> [<https://perma.cc/ZHV8-U5QX>] ("... if such a diagnostic system is designed to take the scan, read it, make the diagnosis, and then present it to the physician who acts merely as a messenger between the system and the patient, then it would seem that the physician is playing a relatively passive role in this provision of treatment").

<sup>77</sup>Curtis E. A. Karnow, *The Application of Traditional Tort Theory to Embodied Machine Intelligence*, in ROBOT LAW 69 (Ryan Calo, A. Michael Froomkin & Ian Kerr eds., 2016).

<sup>78</sup>See, e.g., A. Michael Froomkin, Ian Kerr & Joelle Pineau, *When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning*, 61 ARIZ. L. REV. 33, 51 (2019) (recommending changes to medical malpractice law to reflect AI's superior performance over human physicians).

<sup>79</sup>74 AM. JUR. 2d Torts § 7.

<sup>80</sup>Price, Gerke & Cohen, *supra* note 48, at 1765.

<sup>81</sup>See Phillip G. Peters, Jr., *The Quiet Demise of Deference to Custom: Malpractice Law at the Millennium*, 57 WASH. & LEE L. REV. 163, 165 (2000).

<sup>82</sup>Michael D. Greenberg, *Medical Malpractice and New Devices: Defining an Elusive Standard of Care*, 19 HEALTH MATRIX 423, 428-29 (2009) (noting that half of the states have adopted the reasonable physician standard).

<sup>83</sup>*Id.* at 430 ("... all versions of the malpractice standard are ultimately based on an evaluation of the appropriateness of a physician's conduct, by comparison to what reasonable physicians either do, or should do, in

One obstacle in applying a negligence framework relates to the uncertainty surrounding the standard of care. This is partly explained by the fact that, like any new medical device, the risk implications of clinical AI remain ambiguous. This in turn introduces ambiguity into the standard of care.<sup>84</sup> A related difficulty is determining whether a physician satisfied their duty of care in the absence of any principled basis for a reasonable physician to reject AI recommendations.<sup>85</sup> As discussed above, AI errors may be inherently unforeseeable owing to the opacity of the computational models used to generate decisions or recommendations.<sup>86</sup> That is, the algorithms used in a clinical AI system may be non-transparent because they rely on rules that are too complex for humans to explicitly understand. The opacity can also derive from the fact that no one, not even the programmers, knows what factors go into the machine-learning process that generates the output.<sup>87</sup> For example, a physician will not be able to reject an AI system's recommendation for a personalized medical treatment on the basis of a better counterfactual treatment. The physician cannot refer to generalized medical studies as a basis for their decision since recommendations are personalized for each patient.<sup>88</sup> One can only know *ex post facto* whether it was the right treatment or not. Given that there has not been any case law on medical AI, uncertainty remains as to how a court

---

similar circumstances. The latter is usually determined by reference to the customary practices of other physicians, as established through expert testimony”).

<sup>84</sup>*Id.* at 434. *See also* EUR. COMM'N, *supra* note 4, at 23 (“Emerging digital technologies make it difficult to apply fault-based liability rules, due to the lack of well-established models of proper functioning of these technologies and the possibility of their developing as a result of learning without direct human control”).

<sup>85</sup>*See* Schönberger, *supra* note 46, at 197 (“...in light of the opacity inherent in AI systems, it might indeed be an insurmountable burden for a patient to prove not only causation but the breach of a duty of care in the first place”). Admittedly, explainability is not the only epistemic warrant for following an AI recommendation. *See* I. Glenn Cohen, *Informed Consent and Medical Artificial Intelligence: What to Tell the Patient*, 108 GEO. L. J. 1425, 1443 (2020) (“The epistemic warrant for that proposition [that AI is likely to lead to better decisions] need not be firsthand knowledge – we might think of medical AI/ML as more like a credence good, where the epistemic warrant is trust in someone else”). The challenge lies in identifying the right epistemic warrant or indicia of reliability in the absence of explainability. On this point, some have suggested the need to validate system performance in prospective trials. *See, e.g.,* Alex John London, *Artificial Intelligence and Black-Box Medical Decisions Accuracy versus Explainability*, 49 HASTINGS CTR. REP. 15, 20 (2019). However, owing to variance in organizational factors, a clinical AI system that is found safe and effective in one setting may prove to be significantly less so in another. *See* Sara Gerke et al., *The Need for a System View to Regulate Artificial Intelligence/Machine Learning-Based Software as Medical Device*, 3 NPJ DIGI. MED. 1, 2 (2020) (“due to their systemic aspects, AI/ML-based [software as medical device] will present more variance between performance in the artificial testing environment and in actual practice settings, and thus potentially more risks and less certainty over their benefits”). Nicholson Price has suggested an approach whereby providers would require some validation prior to relying on a black-box algorithm for riskier interventions. This validation would likely come in the form of procedural checks or independent computations by third parties, as opposed to clinical trials. For the riskiest and most counter-intuitive interventions, Price suggests that no black-box verification would be able to overcome the presumption of harm under a reasonable standard of care. He acknowledges however that there may be challenges of implementation, overcaution, and under-compensation associated with this risk-based approach. W. Nicholson Price II, *Medical Malpractice and Black-Box Medicine*, in *BIG DATA, HEALTH LAW, AND BIOETHICS* 301-02 (Cohen et al. eds., 2018).

<sup>86</sup>Admittedly, this is less of a problem in jurisdictions that have kept the customary medical practice standard as the plaintiff would only have to demonstrate that the physician did not do what is customarily done. However, there will likely be a high risk of liability during the transition phase that precedes the emergence of a prevailing customary practice (or one that courts consider dispositive for the purposes of liability). This may also negatively impact technological innovation as “the baseline grounding of the standard of care in customary practice privileges hewing to tradition.” Price II, *supra* note 85, at 304

<sup>87</sup>*Id.* at 299 n.15.

<sup>88</sup>*See* Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B. U. L. REV. 1315, 1338 (2020).

would address a situation involving a divergence between the standard of care and an AI recommendation.<sup>89</sup>

Many of the ambiguities and challenges surrounding the physician standard of care apply equally to negligence claims against hospitals. A hospital can be held liable pursuant to theories of corporate liability and vicarious liability.<sup>90</sup> Under the former, a health care organization can be found directly liable for failure to safeguard patient safety and welfare.<sup>91</sup> Corporate hospital liability's scope encompasses patient injuries sustained as a result of inadequate maintenance of new medical equipment or as a consequence of inadequate policies to ensure that staff have the proper training and expertise.<sup>92</sup> One complication is that a negligence claim against a hospital for harm caused by a clinical AI system would invoke its own standard of care, potentially on the basis of what a reasonable hospital would do in similar circumstances.<sup>93</sup> Uncertainty remains around what courts would recognize as the standard of care for hospitals with respect to the proper use of clinical AI systems.<sup>94</sup> Plaintiffs would also face the potentially significant burden of demonstrating that the hospital had actual or constructive knowledge of the AI system's flaw, or that it failed to train its staff to safely integrate the system into clinical care.<sup>95</sup> Alternatively, a hospital can be found vicariously liable for the negligent acts of its employees.<sup>96</sup> This liability can, in certain cases, extend to the negligent acts of staff physicians, particularly if a hospital imposes workplace rules and regulations on them.<sup>97</sup> That being said, a claim of vicarious liability will be successful only if the plaintiff can establish that the physicians violated their own standard of care.<sup>98</sup> Moreover, "[v]icarious liability claims are particularly challenging in negligence cases, given the diverse contractual relationships that exist between hospitals and

---

<sup>89</sup>This uncertainty is reflected in the scholarship. For instance, Hacker et al. posit that if a physician overrides an ML judgment based on their professional judgement, they can be shielded from liability as negligence "attaches liability only to actions failing the standard of care." Hacker et al., *Explainable AI under Contract and Tort Law: Legal Incentives and Technical Challenges*, 28 *ARTIFICIAL INTELLIGENCE & L.* 415, 424 (2020). In contrast, Cohen et al. think that owing to tort law being inherently conservative, "reliance on medical AI to deviate from the otherwise known standard of care will likely be a defense to liability." Price, Gerke & Cohen, *supra* note 49, at 1766. However, they concede that this may change quickly. *See id.* A recent empirical study has shed some light on how these situations might play out in court. Relying on the results of an online survey of potential jurors, Tobia et al. found that following the standard of care and following the recommendation of AI tools were both effective in reducing lay judgment of liability. Kevin Tobia, Aileen Nielsen & Alexander Stremtizer, *When Does Physician Use of AI Increase Liability?*, 62 *J. NUCLEAR MED.* 17, 20 (2021). In a response to the study, Price et al., acknowledged that, at least with respect to potential jurors and lay knowledge, the use of AI might be close to the standard of care. However, they also note factors that complicate translating the results of the study into real life, including the ability of the judge to resolve cases against patients without trial and the fact that jurors are instructed in the law by the judge and engage in deliberative decision making. As a result, it may be difficult to predict collective juror verdicts from observing only individual juror decisions. W. Nicholson Price II, Sara Gerke & I. Glenn Cohen, *How Much Can Potential Jurors Tell Us About Liability for Medical Artificial Intelligence*, 62 *J. NUCLEAR MED.* 15, 19-20 (2021).

<sup>90</sup>See Sharona Hoffman & Andy Podgurski, *E-Health Hazards: Provider Liability and Electronic Health Record Systems*, 24 *BERKELEY TECH. L.J.*, 1523, 1535 (2009).

<sup>91</sup>*See id.*; *infra* Part IV.C.

<sup>92</sup>Greenberg, *supra* note 82, at 439.

<sup>93</sup>*Id.*

<sup>94</sup>*See e.g.*, Iria Giuffrida & Taylor Treece, *Keeping AI Under Observation: Anticipated Impacts on Physicians' Standard of Care*, 22 *TUL. J. TECH. & INTELL. PROP.* 111, 117-118 (2020).

<sup>95</sup>*See* Efthimios Parasidis, *Clinical Decision Support: Elements of a Sensible Legal Framework*, 20 *J. HEALTH CARE L. & POL'Y* 183, 214 (2018)

<sup>96</sup>*See* Hoffman & Podgurski, *supra* note 90, at 1536 ("The doctrine of 'respondeat superior,' which literally means 'let the superior answer,' establishes that employers are responsible for the acts of their employees in the course of their employment").

<sup>97</sup>*Id.*

<sup>98</sup>Greenberg, *supra* note 82, at 439.

physicians, and factual issues surrounding the level of control exerted by the principal over the agent.”<sup>99</sup>

As for negligence claims against AI manufacturers, the relevant duty of care will likely be whether the manufacturer provided a faulty AI system.<sup>100</sup> A key element to establishing the standard of care for AI manufacturers is to determine whether there exists a custom or usage in the industry applicable to the AI system in question.<sup>101</sup> However, custom establishes only evidence of ordinary care, and the exact standard must be determined on a case-by-case basis.<sup>102</sup> A defendant manufacturer would likely counter by arguing that liability should be premised on the plaintiff establishing that a reasonable manufacturer, taking into account industry custom or usage, would have detected and corrected the defect in the system.<sup>103</sup> This will not be a simple task. As one author has noted:

In determining the scope of a vendor’s duty, relevant factors include the foreseeability of the harm, the connection between the [clinical decision support] recommendation and the harm, the burden on the vendor in structuring alternative recommendations, the feasibility and practicality of affording adequate protections, and the public policy desire to prevent the harm.<sup>104</sup>

Assuming that the standard of care is ascertainable in a given situation, the plaintiff will once again encounter the evidentiary burden of proving that this standard was breached. Meeting this burden will be difficult and expensive.<sup>105</sup>

Ultimately, the decision to use AI for a particular patient is tethered to the decision to use AI in general. The clinician is left in a position where the decision to use AI hinges on an “article of faith” with respect to whether the recommendation will work out for a particular patient.<sup>106</sup> In using black box clinical AI systems, physicians and hospitals “place trust not only in the equation of the model, but also in the entire database used to train it and, in the handling (e.g. labelling) of that database by the designers.”<sup>107</sup> Andrew Selbst has characterized this predicament as a question of whether harms caused by AI systems are sufficiently foreseeable such that a reasonable person can be held liable for their occurrence. In Selbst’s view, negligence law compensates for humans’ “bounded rationality” with the requirement of foreseeability.<sup>108</sup> Because many harmful consequences can be imagined, the law must have some way of deciding at what point an agent has accepted an impermissible amount of risk. Our rationality is bounded by the fact that we have neither perfect information nor the capacity to process all the variable risks involved in a given situation. This limitation means that courts must conduct an inquiry

---

<sup>99</sup>Parasidis, *supra* note 95, at 214.

<sup>100</sup>*Id.*, at 215.

<sup>101</sup>See Michael D. Scott, *Tort Liability for Vendors of Insecure Software: Has the Time Finally Come*, 67 MD. L. REV. 425, 446 (2008).

<sup>102</sup>*Id.*

<sup>103</sup>See Parasidis, *supra* note 94, at 215.

<sup>104</sup>*Id.*

<sup>105</sup>See EUR. COMM’N, *supra* note 4, at 24 (“The more complex the circumstances leading to the victim’s harm are, the harder it is to identify relevant evidence. For example, it can be difficult and costly to identify a bug in a long and complicated software code. In the case of AI, examining the process leading to a specific result (how the input data led to the output data) may be difficult, very time-consuming and expensive”).

<sup>106</sup>Selbst points out that while interpretability can render some AI errors predictable and thus resolve the foreseeability problem, this only works some of the time. Selbst, *supra* note 88, at 1341.

<sup>107</sup>Quinn et al., *supra* note 40.

<sup>108</sup>Selbst, *supra* note 88, at 1360-61. Selbst notes that it “is a fundamental tenet of negligence law that one cannot be liable for circumstances beyond what the reasonable person can account for.” *Id.* at 1360.

into: “(1) what is it reasonable for a person to know?; and 2) [h]ow much can we expect them to be able to process?”<sup>109</sup> Decision-assistance AI systems, with their exponentially greater processing power, exist to compensate for our inherent limitations. However, this does not mean that we have succeeded entirely in “unbounding” our rationality. Rather, AI opacity means that we find ourselves unable to understand and, therefore, supervise or question the decision-making processes. Selbst sees this as being the real challenge to the adoption of negligence law for AI harms:

... it is in precisely the contexts where human limitations currently cause the most injuries that demand for AI will be the greatest. Thus, though the injury rates may improve overall with AI, the people who are injured—and there may still be many—will be without remedy if negligence treats AI errors as functionally unforeseeable.<sup>110</sup>

In short, AI’s opacity will make it a complicated matter for courts to determine the harms that are reasonably foreseeable in the case of clinical AI systems. Even if the harm is somehow foreseeable, it can be burdensome for the plaintiff to prove that there was a breach of the standard of care, especially when the harm in question may have resulted from the interaction of multiple actors.<sup>111</sup> All of this points to negligence being a sub-optimal liability scheme for AI harms. Not only will the law fail to adequately deter potential tortfeasors, but it will also lead to an increase in the costs of adjudicating legal disputes over the standard of care.<sup>112</sup>

### C. AGENCY LAW

More recently, scholars have proposed creative extensions of agency law to address some of the concerns raised above. Matthew Scherer, a proponent of this approach, holds that agency law does not depend on the characteristic of the agent (i.e., it does not need to be a legal person) but rather on the *relationship* between the principal and agent.<sup>113</sup> As such, an AI system could be considered an agent of a principal despite lacking legal personhood.<sup>114</sup> Under this framework, the principal of an AI system would be held vicariously liable for the tortious acts of an AI system when the system “acts within the scope of the agency.”<sup>115</sup> The advantage of an agency approach is in allowing victims to hold the principal(s) liable even where the agent cannot, itself, be held liable.<sup>116</sup> At the same time, agency law contemplates that agents will act autonomously and use their discretion in carrying out the principal’s tasks. That an agent might even use its autonomy

<sup>109</sup>*Id.* at 1361.

<sup>110</sup>*Id.* at 1362-63.

<sup>111</sup>See Price II, Gerke & Cohen, *supra* note 89, at 15-16; George Maliha et al., *Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation*, 99 MILBANK Q. 629, 630 (2021) (noting that “physicians exist as part of an ecosystem that also includes health systems and AI/ML device manufacturers. Physician liability over use of AI/ML is inextricably linked to the liability of these other actors”).

<sup>112</sup>See Rachum-Twaig, *supra* note 53, at 1164.

<sup>113</sup>See Scherer, *supra* note 51, at 286.

<sup>114</sup>The requirements for legal agency are set out in the RESTATEMENT (SECOND) OF AGENCY § 1 (AM. L. INST. 1958):

- (1) Agency is the fiduciary relation which results from the manifestation of consent by one person to another that the other shall act on his behalf and subject to his control, and consent by the other so to act.
- (2) The one for whom action is to be taken is the principal.
- (3) The one who is to act is the agent.

<sup>115</sup>Scherer, *supra* note 51, at 287.

<sup>116</sup>*Id.* at 286.

to contradict the express instructions of the principal expressly falls within the contemplation of agency law.<sup>117</sup>

Who should be considered an AI's principal? Scherer takes a broad approach and argues that an AI system's principals should be the "designers, manufacturers, and developers i.e., those who gave the A.I. system the ability to do legally meaningful things."<sup>118</sup> This would essentially capture anyone involved in the A.I. system's design, manufacturing, updating, maintenance, and use. An AI system could be considered an agent to multiple co-principals or, alternatively, a subagent of a principal. The scope of agency could be "defined in terms of the AI system's capabilities and the precautions that the system's upstream designers and deployers took to prevent downstream operators and users from expanding or altering those capabilities."<sup>119</sup> There are limits to liability. Designers should be excused from liability "if a downstream individual or entity modifies the system in a manner that makes it capable of performing tasks that go beyond even its learnable capabilities."<sup>120</sup> This immunity from liability is qualified by the designer's duty to ensure there are safeguards built into the AI system against potentially dangerous modifications.<sup>121</sup>

It is unclear, however, whether agency law has the conceptual resources to justify holding both health care actors *and* the manufacturer liable for harms caused by clinical AI systems. According to the Restatement (Third) of Agency, a "principal" designates "a person who has authorized another to act on his account and subject to his control."<sup>122</sup> This control does not have to be physical, but the law is clear that the principal must have the *right* to exercise control over the agent's activities.<sup>123</sup> For instance, an airplane owner is not considered the principal of a pilot because the owner had no right to control the pilot's performance during flights.<sup>124</sup> Similarly, the manufacturer of a clinical AI system lacks not only control over how the hospital or the physician actually uses the AI system, but also any *right* to control this use. It is the physician who directly employs the technology in diagnosing and treating patients while the healthcare institution "selects, installs, trains, and operates an AI system that its physicians may utilize."<sup>125</sup> It bears noting that the underlying justification for the theory of vicarious liability is the employer's right to control the means and methods of the employee's work. The threat of being held vicariously liable presumably incentivizes the employer to develop and implement sound procedures to control their employees.<sup>126</sup> In the case of clinical AI systems, any putative line of control between the manufacturer and the AI system is disrupted by the presence and

<sup>117</sup>*Id.* at 289.

<sup>118</sup>*Id.* at 287.

<sup>119</sup>*Id.* at 287-88.

<sup>120</sup>*Id.* at 288.

<sup>121</sup>*Id.*

<sup>122</sup>RESTATEMENT (SECOND) OF AGENCY § 1 cmt. d (AM. L. INST. 1958).

<sup>123</sup>"The person represented has a right to control the actions of the agent." RESTATEMENT (THIRD) OF AGENCY § 1.01 cmt. C (AM. L. INST. 1958). One exception to this rule is the doctrine of apparent agency (i.e., ostensible agency) which holds that a hospital could be liable for an independent contractor's negligence if it represented the contractor as its employee and the patient justifiably relied on the representation. In this case, the hospital is vicariously liable despite having no right of control over the contractor. See Arthur F. Southwick, *Hospital Liability: Two Theories Have been Merged*, 4 J. LEGAL MED. 1, 9-13 (1983).

<sup>124</sup>*Nava v. Truly Nolen Exterminating*, 140 Ariz. 497, 683 P.2d 296, 299-300. Cf. *S. Pac. Transp. v. Cont'l Shippers*, 642 F.2d 236, 238-39 (8th Cir. 1981) (holding that shipper-members of a shippers' association were agents of the association because the association had actual authority to act as the agent for the member defendants and the association was controlled by its members).

<sup>125</sup>Gary E. Marchant & Lucille M. Tournas, *AI Health Care Liability: From Research Trials to Court Trials*, J. HEALTH & LIFE SCI. L. 23, 37 (2019).

<sup>126</sup>Southwick, *supra*, note 123, at 4.



actions of health care intermediaries.<sup>127</sup> This remains the case unless the physician (and perhaps even the health care institution) is entirely taken out of the loop.<sup>128</sup> Until that happens, to include the manufacturer as a principal for the purposes of vicarious liability would conflict with the underlying justification of agency law.

Alternatively, one can take a more selective approach in determining who counts as the principal. In a recent article, Anat Lior proposed an agency law approach to AI harms where the “identity of the principal will change per instance and will heavily depend on the circumstances of the accident.”<sup>129</sup> Under her proposal, the owner and operator of an AI system may be considered principals in one circumstance while the manufacturer may be considered the principal in another.<sup>130</sup> The problem with this approach is that it will often be a contentious matter to determine which party, if any, exercised the relevant control and supervision. Despite being a case on robotics and the LI principle, the Supreme Court of Washington decision in *Taylor v. Intuitive Surgical Inc.*<sup>131</sup> is instructive in this regard. A physician disregarded the robot manufacturer’s guidelines when using the device on a patient, leading to post-surgery complications and death. On appeal, the question facing the court was whether the manufacturer had a duty to warn the hospital in addition to the physician. The decision resulted in a sharp disagreement between the majority and dissenting opinions over the duties of each party in relation to the patient. The majority held that the manufacturer owed duties to the patient—duties that could only be discharged by warning the hospital.<sup>132</sup> The dissent, in contrast, would not have held the manufacturer liable based on there being “several steps” between the manufacturer and the patient.<sup>133</sup> The case of *Taylor* suggests that the connection between a party’s activity and the harmful actions of the clinical AI system will often be highly attenuated, which makes it unwieldy to identify who is properly the principal.<sup>134</sup>

The underlying problem with agency law is that the notion of control or right to control is ultimately too limited to capture many of the material ways in which an actor can be responsible for harm arising from the use of clinical AI systems. AI-induced harms are usually the product of the actions and omissions of multiple actors, with few of these actors exercising direct (or even indirect) control or supervision over the AI system.<sup>135</sup> While the ‘many hands’ problem has been attributed to various AI technologies (notably autonomous

---

<sup>127</sup>On this point, one might note a parallel with the LI doctrine under products liability law whereby the manufacturer is shielded from liability precisely because the product (e.g., a prescription drug) interacts with the plaintiff through a professional intermediary (i.e., the physician).

<sup>128</sup>See Allen, *supra* note 8.

<sup>129</sup>Anat Lior, *AI Entities as AI Agents: Artificial Intelligence Liability and the AI Respondeat Superior Analogy*, 46 MITCHELL HAMLINE L. REV. 1043, 1092 (2020).

<sup>130</sup>*Id.* at 1092-93.

<sup>131</sup>*Taylor v. Intuitive Surgical, Inc.*, 389 P.3d 517, 520 (Wash. 2017).

<sup>132</sup>*Id.*, at 526 (“While doctors are recognized as the gatekeepers between the physician and patient, the hospital is the gatekeeper between the physician and the use of the da Vinci System since the hospital clears surgeons to use it. Thus, the hospital must have warnings about its risks and no tort doctrine should excuse the manufacturer from providing them”).

<sup>133</sup>*Id.*, at 531 (“ISI manufactured the product, ISI sold the product to Harrison, Harrison credentialed the doctor, and the doctor ultimately operated on Taylor’s Husband using the product”).

<sup>134</sup>See Terry, *supra* note 11, at 161-62 (“*Taylor* puts several future issues on display. For example, which members of the distribution chain will face liability and under what legal theory and what are the relative responsibilities of hospitals and developers in training physicians and developing or enforcing protocols for the implementation of AI generally or its use in a particular case?”).

<sup>135</sup>See Mihailis E. Diamantis, *Algorithms Acting Badly: A Solution from Corporate Law*, 89 GEO. WASH. L. REV. 801, 820 (2021) (“The algorithmic misbehavior may result from an unexpected interaction between the algorithm (programmed by one company), the way it is used (by a second company), and the hardware running it (owned by a third company)”).

vehicles),<sup>136</sup> the problem is arguably more acute in the medical context since whether and how a clinical AI system is used will depend on the interactions of numerous actors, processes, and institutions. These include members of the care team, hospital systems, malpractice insurers, and regulators. Other potential elements include the payment structure, data providers, software components providers, and trainers.<sup>137</sup> Few of the interactions between these elements will involve control of the AI system in any meaningful sense of the term. In predicating responsibility exclusively on a principal's (or principals') supposed control or supervision of its agents and subagents, agency law is simply too limited of a legal theory to account for these complex layers of interactions and relationships.<sup>138</sup> As such, an agency law approach may struggle to account for various actors' contribution to making the use clinical AI systems more prone to harmful errors in instances where these actors did not exercise the kind of control that is the touchstone of the principal-agent relation.

#### D. AI PERSONHOOD

Perhaps the most contentious proposed framework for AI liability consists in giving AI systems personhood. In his seminal 1992 article, Lawrence Solum famously suggested that the law could recognize a limited form of legal personhood for AI systems capable of serving as a limited-purpose trustee.<sup>139</sup> More recently, a 2017 report from the European Parliament opened the door to recognizing sophisticated autonomous robots as "having the status of electronic persons responsible for making good any damage they may cause, and possibly applying electronic personality to cases where robots make autonomous decisions or otherwise interact with third parties independently."<sup>140</sup> The authors of the report also recommended the establishment of a compulsory insurance scheme whereby producers or owners of robots would be required to take out insurance to compensate for damages caused by the robots, as well as a compensation fund for damages not covered by the insurance scheme.<sup>141</sup> The proposal was met with vociferous opposition by a group of AI experts who argued that adopting such a status would be ethically and legally inappropriate.<sup>142</sup> This opposition was echoed in a report by a

---

<sup>136</sup>See, e.g., Mark Coeckelbergh, *Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability*, 26 SCI. & ENG'G ETHICS 2051, 2056 (2020).

<sup>137</sup>Gerke et al., *supra* note 85, at 2 (arguing that AI-based medical devices are systems composed of interacting elements and whose performance is dependent on "organizational factors such as resources, staffing, skills, training, culture workflow and processes").

<sup>138</sup>Another indication for why vicarious liability might not be a good fit for clinical AI systems can be found in the decline of the Captain of the Ship doctrine. The idea is that the chief surgeon, as the captain of the ship during surgery, is vicariously liable for the negligence of any person serving on the surgical team. The underlying justification for the doctrine is the right of control one had over the negligent activities of others. However, this justification became increasingly untenable as the size of medical teams grew and as medical professional such as anesthesiologists, nurses, surgical assistants became recognized as performing independent functions. Arthur Southwick draws the following lesson from the decline of this doctrine: "[w]hen medical care is provided by a highly specialized, sophisticated team of professional individuals all working within an institutional setting. It is frequently difficult to determine at any given point in time who is exercising direct control over whom." Southwick, *supra* note 123, at 14-16.

<sup>139</sup>Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L. REV. 1231, 1252-53 (1992).

<sup>140</sup>COMM. ON LEGAL AFFAIRS, EUR. UNION PARLIAMENT, REP. WITH RECOMMENDATIONS TO THE COMM'N ON CIV. L. RULES ON ROBOTICS, at 18 (2017).

<sup>141</sup>*Id.* at 17-18.

<sup>142</sup>A.I. and Robotics Experts, *Robotics Open Letter to the European Commission* 1 (Apr. 5, 2018), <https://g8fip1kplyr33r3krz5b97d1-wpengine.netdna-ssl.com/wp-content/uploads/2018/04/RoboticsOpenLetter.pdf> [<https://perma.cc/45G8-YB7L>].

European Commission expert group tasked with examining the question of liability for artificial intelligence.<sup>143</sup>

Despite this high-profile opposition, there remains a lively debate among scholars over the merits of AI personhood. Some of those who support the idea focus on autonomous AI systems as being *analogous* to natural persons.<sup>144</sup> There have been arguments, for instance, for IBM's WFO to be considered legally analogous to a consulting physician<sup>145</sup> or medical student.<sup>146</sup> Some commentators have even suggested granting fully autonomous AI systems state licensure for the practice of medicine.<sup>147</sup> Others emphasize the idea of AI personhood as an instrument to achieve socially useful ends.<sup>148</sup> Along these lines, some have argued for the legitimacy of robot personhood on the basis that corporations have long been accorded personhood status despite lacking key features of natural persons.<sup>149</sup>

While it is beyond the scope of this Article to resolve the complex question of AI personhood, certain points are worth bearing in mind. First, it seems risky in the near term to accord AI full-blown legal personhood with rights and duties on par with natural persons given the host of ethical concerns that have yet to be addressed.<sup>150</sup> Second, while the recognition of limited AI personhood (e.g., analogous to corporate personhood) may not be ethically problematic to the same degree as full-blown personhood, careful consideration should be given to the specific purpose that such recognition would serve and what

<sup>143</sup>EUR. COMM'N, *supra* note 4, at 38 ("Harm caused by even fully autonomous technologies is generally reducible to risks attributable to natural persons or existing categories of legal persons, and where this is not the case, new laws directed at individuals are a better response than creating a new category of legal person").

<sup>144</sup>*See, e.g.,* Nadia Banteka, *Artificially Intelligent Persons*, 58 HOUS. L. REV. 537, 552 (2021) ("Even amongst established legal persons such as human beings, legal systems have created categories of humans with more or less rights and different sets of obligations. Consider, for instance, the rights enjoyed by an adult human to those enjoyed by a child. By analogy, artificial entities also fall on this spectrum and *have often been conferred legal personhood* with more or less restricted bundles of rights and obligations") (emphases added); Russ Pearlman, *Recognizing Artificial Intelligence As Authors and Investors Under U.S. Intellectual Property Law*, 24 RICH. J.L. & TECH. 1, 29 (2018) (suggesting that AI should be granted analogous legal personhood, such as that granted to corporations and government entities); Ryan Abbott & Alex Sarch, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, 53 U.C. DAVIS L. REV. 323, 356 (2019) (suggesting ways in which AI criminality should be considered analogously to natural persons' criminality).

<sup>145</sup>Allain, *supra* note 74, at 1062-63.

<sup>146</sup>Jason Chung & Amanda Zink, *Hey Watson – Can I Sue You for Malpractice? Examining the Liability of Artificial Intelligence in Medicine*, 11 ASIA PAC. J. HEALTH L. & ETHICS 51, 53 (2018).

<sup>147</sup>*See, e.g.,* Allen, *supra* note 8, at 1177-78.

<sup>148</sup>WHITE & CHOPRA, *supra* note 63, at 158 ("to ascribe legal personhood to an entity is to do no more than to make arrangements that facilitate a particular set of social, economic and legal relationships").

<sup>149</sup>*See, e.g.,* JACOB TURNER, *ROBOT RULES: REGULATING ARTIFICIAL INTELLIGENCE* 184-188 (2019) (arguing that "[g]ranted AI legal personality could be a valuable firewall between existing legal persons and the harm which AI could abuse. Individual AI engineers and designers might be indemnified by their employers, but eventually creators of AI systems – even at the level of major corporates – may become increasingly hesitant in releasing innovative products to the market if the programmers are unsure as to what their liability will be for unforeseeable harm"). *Cf.* Vikram R. Bhargava & Manuel Velasquez, *Is Corporate Responsibility Relevant to Artificial Intelligence Responsibility?* 17 GEO. J. L. & PUB. POL'Y 829, 829, 833 (arguing that the reasons for holding corporations responsible are inapplicable to AI agents since corporations are made up of and act through agents, which is not the case for AI).

<sup>150</sup>*See, e.g.,* Ugo Pagallo, *Vital, Sophia, and Co. – The Quest for the Legal Personhood of Robots*, 9 INFORMATION 230, 236-37 (2018) (noting that artificial agents lack self-consciousness, human-like intentions, and the ability to suffer – the requisites associated with granting someone, or something, legal personhood); Ryan Abbott & Alex Sarch, *Punishing Artificial Intelligence: Legal Fiction or Science Fiction*, 53 U.C. DAVIS L. REV. 103, 154 (2019) (holding that "[f]ull-fledged legal personality for AI is equivalent to that afforded to natural persons, with all the legal rights that natural persons enjoy, would clearly be inappropriate"); John-Stewart Gordon, *Artificial Moral and Legal Personhood*, 36 A.I. & Soc'Y 457, 470 (2021) (arguing that artificially intelligent robots currently fail to meet the criteria of rationality, autonomy, understanding, and social relations necessary for moral personhood).

advantage it has, if any, over other legal solutions.<sup>151</sup> Third, even limited AI personhood will require robust safeguards such as having funds or assets assigned to the AI person.<sup>152</sup> Fourth, assuming that such safeguards are in place, it is conceivable that recognizing limited AI personhood could serve useful cost-spreading and accountability functions.<sup>153</sup> On this point, careful consideration will also have to be given to how AI personhood fits into and reinforces existing liability frameworks.

#### E. THEORIES OF ENTERPRISE LIABILITY

EL holds that losses caused by an enterprise should be borne by the enterprise or the activity itself.<sup>154</sup> Unlike agency law, a direct relationship of control is not a prerequisite to a finding of liability; the focus is rather on distributing enterprise's accident costs broadly among members of the enterprise. EL is premised on the principle that "[t]he costs of an injury should be shared by those who profit from the activity responsible for the injury; they should not be concentrated on the injured party or be dispersed across unrelated activities."<sup>155</sup> Similar to products liability, fault is not an element of EL.<sup>156</sup>

The idea of applying some form of EL to health care has been the subject of intense legal scholarship going back decades.<sup>157</sup> In 1991, the American Law Institute ("ALI") published a report that proposed shifting the locus of liability from individual physicians to the health care institution.<sup>158</sup> Under this proposal, physicians are exculpated from liability (eliminating their need to purchase liability insurance) on the condition that the health care institution takes out insurance to allow the patient to recover for injuries caused by the physician.<sup>159</sup> Physicians affiliated with a hospital would be treated as members of a single enterprise engaged in delivering health care to patients.<sup>160</sup> This approach was predicated on the belief that the health care organization is in the best position to identify and address the inadvertent mishaps of individual physicians.<sup>161</sup> Paul Weiler, one of the lead authors of the ALI report, advocated for hospital EL on the basis that EL would be a more sensible compensation scheme, more economically efficient, and more effective in preventing injuries to patients.<sup>162</sup> However, the idea of

---

<sup>151</sup>See, e.g., Bert-Jaap Koops, Mireille Hildebrandt & David-Oliver Jaquet-Chiffelle, *Bridging the Accountability Gap: Rights for New Entities in the Information Society?*, 11 MINN. J.L. SCI. & TECH. 497, 560 (2010) (proposing that we consider "whether the attribution of a restricted legal personhood, involving certain civil rights and duties, has added value in comparison with other legal solutions").

<sup>152</sup>See EUR. COMM'N, *supra* note 4, at 38.

<sup>153</sup>See *infra*, Part III.E.

<sup>154</sup>Howard C. Klemme, *The Enterprise Liability Theory of Torts*, 47 U. COLO. L. REV. 153, 158 (1975).

<sup>155</sup>Gregory C. Keating, *The Theory of Enterprise Liability and Common Law Strict Liability*, 54 VAND. L. REV. 1285, 1286 (2001). This theory of liability originated in worker compensation schemes enacted in England and the United States in the early 20th century and has since exerted an influence in various areas of tort law, including products liability law.

<sup>156</sup>*Id.* at 1287-88.

<sup>157</sup>See George L. Priest, *The Invention of Enterprise Liability: A Critical History of the Intellectual Foundations of Modern Tort Law*, 14 J. LEGAL STUD. 461 (1985) for one of the first extended scholarly treatments of the topic.

<sup>158</sup>AM. L. INST., *Medical Malpractice, REPORTERS' STUDY II: ENTERPRISE RESPONSIBILITY FOR PERSONAL INJURY* 111, 113 (1991).

<sup>159</sup>*Id.* at 114-15.

<sup>160</sup>*Id.* at 118.

<sup>161</sup>*Id.* at 123 ("The collective wisdom of the hospital team can be pooled to devise feasible procedures and technologies for guarding against the ever-present risk of occasional human failure by even the best doctors").

<sup>162</sup>See Kenneth S. Abraham & Paul C. Weiler, *Enterprise Medical Liability and the Evolution of the American Health Care System*, 108 HARV. L. REV. 381, 381 (1994).

imposing hospital EL never took off due to worries on the part of health care organizations and lawmakers about the potential costs of a no-fault liability scheme and concerns from physicians about the loss of professional autonomy.<sup>163</sup> There have nonetheless been notable examples of *voluntary* EL.<sup>164</sup>

Other scholars have proposed imposing a system of EL on managed care organizations (“MCOs”).<sup>165</sup> William Sage argued that doing so would result in improved quality, improved compensation for negligence injury, and lower administrative costs.<sup>166</sup> Shifting some accountability away from physicians to MCOs, Sage argued, would track the increasing control that organizations exercise over the provision of health care (e.g., in managing the flow of information between the organization and patients).<sup>167</sup> In contrast to traditional theories of medical malpractice, “enterprise liability explicitly acknowledge[s] that health care has become more an institutional process than a series of discrete interactions between patients and individual physicians.”<sup>168</sup> Importantly, he points to “increasing evidence that most errors in health care delivery, while human in proximate cause, are ultimately the result of faulty institutional processes.”<sup>169</sup> Along a similar line of reasoning, another author has argued that Health Management Organizations (“HMOs”) dictate the parameters of medical care and are locked in a single enterprise affecting patient care.<sup>170</sup> EL has also been suggested for Accountable Care Organizations that were formed in response to the Affordable Care Act.<sup>171</sup>

There have moreover been proposals to extend EL to harms caused by novel medical technology. Thomas McLean has advocated for a single health care provider, specifically the medical service payor, to be liable for negligent acts that occur in the process of performing cybersurgery.<sup>172</sup> The idea is that litigation would be simplified if only one party is held responsible for providing all compensation.<sup>173</sup> As a result, EL would avoid finger-pointing, facilitate litigation, and generally decrease the transaction costs associated with adverse cybersurgical events. EL would also facilitate patient safety by

---

<sup>163</sup> See Phillip G. Peters, Jr., *Resuscitating Hospital Enterprise Liability*, 73 MO. L. REV. 369, 376 (2008); Robert A. Berenson & Randall R. Bovbjerg, *Enterprise Liability in the Twenty-First Century*, in *MEDICAL MALPRACTICE AND THE U.S. HEALTH CARE SYSTEM* 230-232 (William M. Sage & Rogan Kersh eds., 2006). It should be noted EL does not require no-fault liability. See e.g., Vladeck, *supra* note 5, at 147 n.91 (distinguishing a no-fault liability system of EL that imposes mandatory insurance and eliminates access to the judicial system from a strict liability version implemented by the courts).

<sup>164</sup> Some managed care organizations and government organizations such as the VA voluntarily assume liability for the negligent acts of staff. See Daniel P. Kessler, *Evaluating the Medical Malpractice System and Options for Reform*, J. ECON. PERSPS. 93, 102 (2011).

<sup>165</sup> William M. Sage, *Enterprise Liability and the Emerging Managed Health Care System*, 60 L. & CONTEMP. PROBS. 159, 159 (1997).

<sup>166</sup> *Id.* at 166-69.

<sup>167</sup> *Id.* at 167.

<sup>168</sup> *Id.* at 169.

<sup>169</sup> *Id.* at 195.

<sup>170</sup> Jack K. Kilcullen, *Groping for the Reins: ERISA, HMO Malpractice, and Enterprise Liability*, 22 AM. J.L. & MED. 7, 10 (1996). Kilcullen’s proposal was made at a time when HMOs exercised greater control over patients’ health care utilization, such as the choice of providers and hospitals. This control began to loosen in the second half of the 1990s due to consumer and provider backlash. See Ronald Lagoe et al., *Current and Future Developments in Managed Care in the United States and Implications for Europe*, 3 HEALTH RSCH. POL’Y & SYS. 3-4 (2005), <https://health-policy-systems.biomedcentral.com/articles/10.1186/1478-4505-3-4> [<https://perma.cc/C75X-HSET>].

<sup>171</sup> See Laura D. Hermer, *Aligning Incentives in Accountable Care Organizations: The Role of Medical Malpractice Reform*, 17 J. HEALTH CARE L. & POL’Y 271, 273 (2014).

<sup>172</sup> Thomas R. McLean, *Cybersurgery – An Argument for Enterprise Liability*, 23 J. LEGAL MED. 167, 207 (2002). The medical service payor could be “the federal government or Fortune 500 insurance company doing business as a managed care organization.” *Id.* at 208.

<sup>173</sup> *Id.* at 207.

financially incentivizing medical service providers to choose only the best technology and conduit service providers.<sup>174</sup> More recently, EL has been suggested as a liability framework for harm caused by medical AI systems, though there have been few attempts to spell out such a framework in any great detail.<sup>175</sup>

A further development of EL, and one that has been recently proposed for AI harms, is CEL. Under the settled or classical version of the CEL doctrine, “entities within a set of interrelated companies may be held jointly and severally liable for the actions of other entities that are part of the group.”<sup>176</sup> David Vladeck has proposed a variation of classical CEL as a response to the question of who should bear the cost of harms caused by autonomous vehicles under a strict liability regime.<sup>177</sup> Here, he draws inspiration from a line of federal cases involving deceptive marketing practices where the theory was used to hold a group of corporate entities liable for harm directly caused by only one member of the group. In the leading CEL case of *FTC v. Tax Club Inc.*, the U.S. District Court for the Southern District of New York affirmed CEL as an exception to the rule against group pleading, i.e., lumping defendants together in a way that does not distinguish the misconduct of each.<sup>178</sup> The CEL exception has been applied to cases of corporate misconduct where defendants strategically formed and used various corporate entities to violate consumer protection law.<sup>179</sup> These corporations were considered to be functioning jointly as a common enterprise for the purposes of liability.<sup>180</sup>

In Vladeck’s variation of CEL, it suffices that legal entities work toward a common end.<sup>181</sup> Applying this principle to the situation of AI, component manufacturers are considered to be engaged in the common objective of designing, programming, and manufacturing a vehicle despite not functioning jointly. The reasoning why CEL is appropriate in the context of AI is that:

A common enterprise theory permits the law to impose joint liability without having to lay bare and grapple with the details of assigning every aspect of wrongdoing to one party or another; it is enough that in pursuit of a common aim the parties engaged in wrongdoing. That principle could be engrafted onto a new, strict liability regime to address

---

<sup>174</sup>*Id.*

<sup>175</sup>*See, e.g.*, Allen, *supra* note 8, at 1177 (noting that EL’s removal of the need to prove negligence may help manage the risk of patient harm from AI). Jessica Allain has proposed a statutory scheme whereby an action against an AI system like Watson could proceed under EL, with the enterprise consisting of the AI as a legal person, the AI’s owner, and the physicians involved. However, EL would only be triggered once a panel of experts has determined to the court’s satisfaction that there was no hardware failure; otherwise, the action would proceed under products liability. *See* Allain, *supra* note 74, at 1076-1077. The concern here is that this preliminary step of assessing hardware failure risks being time- and resource-intensive, which would inject an additional layer of uncertainty to the recovery process.

<sup>176</sup>Vladeck, *supra* note 5, at 149.

<sup>177</sup>*Id.*

<sup>178</sup>*Fed. Trade Comm’n v. Tax Club, Inc.*, 994 F. Supp. 2d 461, 469 (S.D.N.Y. 2014); *see also* *Consumer Fin. Prot. Bureau v. NDG Fin. Corp.*, No. 15-cv-5211, 2016 WL 7188792, at \*16 (S.D.N.Y. Dec. 2, 2016); *Fed. Trade Comm’n v. 4 Star Resol., LLC*, No. 15-CV-112S, 2015 WL 7431404, at \*1 (W.D.N.Y. Nov. 23, 2015); *Fed. Trade Comm’n v. Vantage Point Servs., LLC*, No. 15-CV-006S, 2015 WL 2354473, at \*3 (W.D. N.Y. May 15, 2015) (specifying that “a common enterprise analysis is neither an alter ego inquiry nor an issue of corporate veil piercing; instead, the entities within the enterprise may be separate and distinct corporations”); *Fed. Trade Comm’n v. Nudge, LLC*, 430 F. Supp. 3d 1230, 1234 (D. Utah 2019); *Fed. Trade Comm’n v. Fed. Check Processing, Inc.*, No. 12-CV-122-WMS-MJR, 2016 WL 5956073, at \*2 (W.D.N.Y. Apr. 13, 2016).

<sup>179</sup>*E.g.*, *Fed. Trade Comm’n v. Pointbreak Media, LLC*, 376 F. Supp. 3d 1257, 1287 (S.D. Fla. 2019).

<sup>180</sup>*Delaware Watch Co. v. Fed. Trade Comm’n*, 332 F.2d 745, 746 (2d Cir. 1964) (holding that the individuals were “transacting an integrated business through a maze of interrelated companies”).

<sup>181</sup>Vladeck, *supra* note 5, at 149.

the harms that may be visited on humans by intelligent autonomous machines when it is impossible or impracticable to assign fault to a specific person.<sup>182</sup>

For Vladeck, the common enterprise is between the manufacturer of the autonomous vehicle and sub-component manufacturers (e.g., of radar and laser sensors, computers). He provides two reasons why the manufacturer should not absorb the full cost of accidents as is often the case under prevailing products law: (1) the component may be at the root cause of the accident and (2) insulating the component manufacturers from liability would not incentivize them from innovating and improving their products.<sup>183</sup>

Vladeck's modified CEL also differs from versions of EL where the cost of accidents is spread among companies engaged in the same hazardous industry.<sup>184</sup> Whereas this kind of industry-wide cost spreading might be appropriate for small and highly concentrated industries, it is less so for expansive industries with many players using different technology and manufacturing processes.<sup>185</sup> Notwithstanding key conceptual differences, there is a strong "family resemblance" between EL, classical CEL, and Vladeck's modified CEL. All of these theories shift away from negligence's tendency to treat accidents as resulting from the misconduct/omission of a single defendant. As such, they allow courts to overcome the finger-pointing problem that arises in situations where responsibility may be dispersed among various actors.<sup>186</sup>

#### IV. COMMON ENTERPRISE LIABILITY: A LIABILITY FRAMEWORK FOR HEALTH AI

Of all the competing liability frameworks, a version of EL is the most appropriate to address harms arising from the use of clinical AI systems. EL's shift from focusing liability on *acts* to *activities* resonates with the dispersion of responsibility among various actors (and networks of actors) involved in the operation of a clinical AI system. To this end, Vladeck's CEL-based theory of liability can be fruitfully applied to the case of health AI systems with one major variation: the law should recognize a common enterprise among the physician, the manufacturer, and the hospital. By appropriating the criteria of "common objective" to determine who is part of a common enterprise, the proposed framework can facilitate the apportioning of responsibilities and liability among disparate actors under a single legal theory.<sup>187</sup>

##### A. CEL AND THE RESPONSIBILITY GAP

As discussed above, injuries caused by health technology are usually the result of the actions and omissions of multiple actors who relate to and influence each other in complex ways. The numerous stakeholders involved in the implementation and operation

<sup>182</sup>Vladeck, *supra* note 5, at 149.

<sup>183</sup>*Id.* at 148.

<sup>184</sup>*Id.* at 129 n.39.

<sup>185</sup>If this does not turn out to be the case, Vladeck concedes that enterprise liability may be more appropriate. *Id.* ("Of course, if the number of driver-less vehicles was relatively small and there were issues of identifying the manufacturer of a vehicle that caused significant harm, enterprise theory of liability might be viable in that situation as well").

<sup>186</sup>*Id.* at 149.

<sup>187</sup>See Andrea Bertolini, Comm. on Legal Affairs, *Artificial Intelligence and Civil Liability*, at 111 (2020), [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL\\_STU\(2020\)621926\\_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf) [<https://perma.cc/G4H6-M9XH>] (holding the importance of apportioning liability between the medical practitioner, AI manufacturer, and hospital/structure that operates the AI system or employs the practitioner).

of a clinical AI system obscure the attribution of fault and therefore responsibility. As Taddeo and Floridi explain:

The effects of decisions or actions based on AI are often the result of countless interactions among many actors, including designers, developers, users, software and hardware. This is known as distributed agency. With distributed agency comes distributed responsibility. Existing ethical frameworks address individual, human responsibility, with the goal of allocating punishment or reward based on the actions and intentions of an individual. They were not developed to deal with distributed responsibility.<sup>188</sup>

Distributed agency is particularly acute in the medical space. While a clinical AI system might generate a wrong treatment recommendation based on a faulty algorithm developed by the manufacturer, it is the physician who makes the final decision. Moreover, the actions of other actors may bear on the physician's decision to endorse the AI recommendation, such as a hospital pressuring its physicians to rely on the AI system's outputs. This culminates in "a situation where each of the stakeholders involved have contributed to medical treatment, with [none] of them being fully to blame."<sup>189</sup>

The result is a responsibility gap whereby it is difficult to assign blame to any one party.<sup>190</sup> This difficulty has prompted calls for a more distributed or collective conception of responsibility with respect to harm caused by clinical AI.<sup>191</sup> Vladek's CEL is a promising solution to the problem of distributed agency in that it allows the allocation of legal responsibility among multiple actors without having to parse out the contributions, interactions, and wrongdoing of each individual actor. This avoids having to investigate out the "long causal chain of human agency" that characterizes the use and development of AI technology—a task that would invariably be challenging and resource intensive.<sup>192</sup>

Agency law is, admittedly, not entirely bereft of conceptual resources to apportion liability to multiple actors on the basis of these actors being co-principals. The difficulty is that, to qualify as an AI principal and thereby be liable, an actor must have the ability right to control the AI system.<sup>193</sup> This would likely shield the manufacturer, and possibly the hospital, from liability given the gaps in control between these parties and the operations of the AI system.<sup>194</sup> For a victim to recover from the enterprise under the proposed CEL-based approach, it needs only be demonstrated that the actors worked toward a common end; there is no requirement that one party had the ability or right to control the actions of the AI system. This approach is, therefore, more responsive to the problem of the responsibility gap and will make it easier for victims to seek compensation.

---

<sup>188</sup>Mariarosaria Taddeo & Luciano Floridi, *How AI Can be a Force for Good*, 361 SCIENCE 751, 751 (2018). Curtis Karnow, in an article published almost 25 years ago, recognized the difficulty of assigning liability to a single actor in situations involving AI harms given the "distributed computing environment in which [artificial intelligence] programs operate." Curtis E. A. Karnow, *Liability for Distributed Artificial Intelligences*, 11 BERKELEY TECH. L. J. 147, 155 (1996).

<sup>189</sup>See Grote & Berens, *supra* note 24, at 209.

<sup>190</sup>See Andreas Matthias, *The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata*, 6 ETHICS & INFO. TECH. 175, 177 (2004) (arguing that "there is an increasing class of machine actions, where the traditional ways of responsibility ascription are not compatible with our sense of justice and the moral framework of society because nobody has enough control over the machine's actions to be able to assume the responsibility for them. These cases constitute what we will call the *responsibility gap*").

<sup>191</sup>See Grote & Berens, *supra* note 24, at 209.

<sup>192</sup>Coeckelbergh, *supra* note 136, at 2057.

<sup>193</sup>See *supra* notes 122-123 and accompanying text.

<sup>194</sup>See *supra* notes 131-138 and accompanying text.



## B. COMMON ENTERPRISE AMONG PHYSICIANS, AI MANUFACTURERS, AND HOSPITALS

While somewhat removed from Vladeck's initial intentions, the idea that the physician, manufacturer, and hospital work toward a common objective is a logical application and extension of CEL theory. It bears noting that clinical AI is conceived and designed to be used by medical health professionals in health care organizations for the purpose of providing care to patients.<sup>195</sup> The pursuit of this common aim is also underlined by AI systems' increasing embeddedness in crucial aspects of health care operations. AI systems have the "potential to facilitate diagnostics, decision-making, big-data analytics, and administration."<sup>196</sup> As such, AI systems are beginning to assume roles that have been traditionally occupied by health care professionals<sup>197</sup> and are already being envisioned as future replacement for some of these positions.<sup>198</sup>

That clinical AI systems are designed to duplicate, complement, or (in certain instances) take over certain defined activities within health care sets them apart from AI systems with more open-ended applications. Clinical AI systems have been designed to address deficiencies that are specific to the practice of medicine and to the operations of health care organizations, deficiencies that have long evaded resolution by human intelligence alone. Indeed, this understanding of health AI as fulfilling compensatory and enhancement functions was explicitly adopted by the American Medical Association ("AMA") in a policy statement expressing preference for the term "augmented intelligence" over artificial intelligence.<sup>199</sup> This terminology was intended to reflect "the enhanced capabilities of human clinical decision making when coupled with these computational methods and systems [for data analysis]."<sup>200</sup> While other medical technologies also fulfill compensatory and enhancement functions to a certain extent, clinical AI systems are distinguished by, among other things, their potential to be ubiquitous in medical interactions and to issue treatment recommendations.<sup>201</sup>

Indeed, Eric Topol observes a "convergence of human and artificial intelligence" at various levels of medicine.<sup>202</sup> At the clinical level, Topol predicts that at some point in the future every clinician will use AI technology involving deep neural networks to help "interpret medical scans, pathology slides, skin lesions, retinal images electro cardiograms, endoscopy, faces, and vital signs."<sup>203</sup> The point here is not so much that AI has fully lived up to its promise or will in the future; rather, it is that manufacturers design health AI systems to augment, enhance, or compensate for the capacities of health care professionals in pursuit of the same objectives.<sup>204</sup> The goal, as Topol notes, is not to develop fully

---

<sup>195</sup>See, e.g., IBM, *Watson Health: Get the Facts* (last visited June 29, 2020), <https://www.ibm.com/watson-health/about/get-the-facts> [<https://perma.cc/LX3W-DBFQ>] ("By combining human experts with augmented intelligence, IBM Watson Health helps health professionals and researchers around the world translate data and knowledge into insights to make more-informed decisions about care in hundreds of hospitals and health organizations").

<sup>196</sup>Bertalan Mesko, Gergely Hetényi & Zsuzsanna Györfy, *Will Artificial Intelligence Solve the Human Resource Crisis in Healthcare?*, 18 BMC HEALTH SERVS. RSCH. 545, 545 (2018).

<sup>197</sup>Allain, *supra* note 74, at 1062.

<sup>198</sup>Mesko, Hetényi & Györfy, *supra* note 196.

<sup>199</sup>AM. MED. ASS'N, AUGMENTED INTELLIGENCE IN HEALTH CARE 2 (2018), <https://www.ama-assn.org/system/files/2019-01/augmented-intelligence-policy-report.pdf> [<https://perma.cc/CC4S-AD9C>].

<sup>200</sup>*Id.*

<sup>201</sup>See Gerke et al., *supra* note 85.

<sup>202</sup>Eric J. Topol, *High-Performance Medicine: The Convergence of Human and Artificial Intelligence*, 25 NATURE MED. 44, 44 (2019).

<sup>203</sup>*Id.*

<sup>204</sup>See Parasidis, *supra* note 95, at 186 (noting that "[t]he goals underlying use of [clinical decision support] mirror those of clinical practice guidelines").

automated AI with no backup clinicians but to achieve “synergy, offsetting functions that machines do best combined with those that are best suited for clinicians.”<sup>205</sup> And while it is physicians that typically operate these systems, they do so in a tightly integrated ecosystem that includes the hospital and AI device manufacturers.<sup>206</sup> All of this suggests that physicians, AI manufacturers, and hospitals are engaged in what can be broadly characterized as a common enterprise.

### C. POLICY REASONS FOR INCLUSION IN THE COMMON ENTERPRISE

There are compelling policy reasons for including the foregoing actors in the common enterprise and thereby holding them liable for harms caused by clinical AI systems. First, the inclusion of the physician is warranted given the physician’s role in operating the AI system, interpreting the output, and endorsing or rejecting the AI system’s recommendation. In a straightforward sense, the physician is the actor most closely implicated in the harm. This seems to be a *prima facie* reason to include the physician as part of the common enterprise, notwithstanding any good faith reliance on the AI’s recommendation or the involvement of other actors. As Maliha and colleagues have observed, courts have allowed malpractice suits to proceed against health professionals even when there were mistakes in medical literature given to patients or when a pharmaceutical company had provided inadequate warning of a therapy’s adverse effects. Moreover, courts have held physicians liable for malpractice based on errors made by system technicians or manufacturers.<sup>207</sup>

The issue becomes complicated given the “black box” nature of AI systems in that a physician will be unable to understand, let alone challenge, the underlying reasoning of the AI recommendation. That being said, it is the physician who ultimately makes an independent judgment whether or not to follow an AI’s recommendation in a particular case.<sup>208</sup> If a physician becomes aware of a malfunction or defect in the AI system, for instance, the physician arguably has a duty to cease use of the equipment and report the problem to the hospital and possibly the manufacturer.<sup>209</sup> It would therefore be questionable policy to exclude the physician from the common enterprise, and thereby allow physicians to rely on (or disregard) AI recommendations with impunity.<sup>210</sup>

Second, the manufacturer is an equally natural candidate for inclusion in the common enterprise despite having little control over the operation of the clinical AI system. Manufacturers have intimate knowledge of the characteristics and features of their products. They are the party that typically exercises control over the product’s design and programming.<sup>211</sup> Accordingly, they are in a unique position to invest in and implement ways to make the AI system safer for end-users such as health care organizations and physicians. As expressed by the European Commission, “it is the producer [of AI systems] who is the cheapest cost avoider and who is primarily in a position to control the risk of accidents.”<sup>212</sup> Moreover, making the manufacturer bear financial responsibility for

<sup>205</sup>Topol, *supra* note 202, at 51.

<sup>206</sup>See Maliha et al., *supra* note 111.

<sup>207</sup>*Id.* at 632.

<sup>208</sup>*Id.* (“Physicians have a duty to independently apply the standard of care for their field, regardless of an AI/ML algorithm output”).

<sup>209</sup>See Giuffrida & Treece, *supra* note 94, at 120.

<sup>210</sup>Moreover, keeping physicians in the enterprise may be justified on the basis that, like the manufacturer and hospital, physicians benefit from the use of AI systems. As discussed below, the internal morality of strict and EL would hold that it is fair to hold the physician at least *prima facie* liable.

<sup>211</sup>See Rachum-Twaig, *supra* note 53, at 1172.

<sup>212</sup>EUR. COMM’N, *supra* note 4, at 40.

injuries caused by their products will incentivize them to research ways to avoid losses that are currently unavoidable.<sup>213</sup>

Finally, the choice to include the hospital may be less obvious given that the hospital neither designs, manufactures, nor directly operates the clinical AI system. As such, one might argue that the hospital is somewhat removed from the harm caused by clinical AI systems. Its inclusion is nonetheless warranted on the basis that hospitals constitute the “major *institutional* bodies responsible for the quality of health care.”<sup>214</sup> The hospital is uniquely situated to guard against the omnipresent risk of dangerous error using the collective wisdom and experience of its members.<sup>215</sup> As the Institute of Medicine observed in its seminal report *To Err is Human* over 20 years ago, medical errors are often the result of faulty systems within health care organizations.<sup>216</sup> This report was prescient in anticipating the human-machine interface as an important focus of preventative efforts.<sup>217</sup> The introduction of new technology invariably raises the possibility of new errors. It falls on the health care organizations to take preventative measures as “safe equipment design and use depend on a chain of involvement and commitment that begins with the manufacture and continues with careful attention to the vulnerabilities of a new device or system.”<sup>218</sup> Among health care actors, hospitals hold arguably the most influence over the kind of technology that is used, how it is used, and by whom it is used.<sup>219</sup>

Including the hospital in the common enterprise should not be taken as an ascription of omniscience or omnipotence to the hospital as an institution. Hospitals are ultimately run by human administrators who can no more predict or prevent any specific AI-induced harm than physicians or (in some instances) AI engineers. Nonetheless, a hospital is well-situated to assess whether a physician has the requisite training, experience, and safety record to treat patients within that hospital’s premises.<sup>220</sup> Where there have been issues of misconduct or malpractice on the part of a certain physician, the hospital is in a unique position to ensure that the physician practices in a way that minimizes risk to patient safety.<sup>221</sup> Similarly, the hospital is well-situated to implement structures and systems to minimize the risks associated with the use of clinical AI systems.<sup>222</sup> Cases involving cybersurgery misadventures such as *Taylor* have revealed the extent to which the introduction of advance technology has intensified the institutional and systems-based character of modern medical practice. Assigning liability to health care organizations could facilitate systemwide improvements in the safe use of medical technology.

---

<sup>213</sup>See Kenneth S. Abraham & Robert L. Rabin, *Automated Vehicles and Manufacturer Responsibility for Accidents: A New Legal Regime for a New Era*, 105 VA. L. REV. 127, 153-54 (2019) (writing in the context of automated vehicles).

<sup>214</sup>Abraham & Weiler, *supra* note 162, at 415 (emphasis added).

<sup>215</sup>AM L. INST., *supra* note 158, at 123.

<sup>216</sup>INST. MED., *supra* note 6, at 169.

<sup>217</sup>*Id.* at 175.

<sup>218</sup>*Id.*

<sup>219</sup>Abraham & Weiler, *supra* note 162, at 416 (noting that “[h]ospitals are typically responsible for selecting and providing the supplies, facilities, and equipment used in treatment, as well as for hiring and firing the employees who play an important role on the patient care team. Hospitals can also grant admitting privileges to physicians, and can restrict, suspend, or terminate the privileges of doctors whose poor quality of treatment has come to the hospital’s attention”).

<sup>220</sup>*Id.*

<sup>221</sup>On this point, Allen suggests that physician medical societies, such as the College of American Pathologists, could inform standards and strategies for managing patient risk in AI implementation. See Allen, *supra* note 8, at 1177.

<sup>222</sup>For instance, the hospital is in a unique position to establish nonpunitive systems for reporting and analyzing AI errors, anticipate errors by double checking for vulnerabilities, train novice practitioners through simulations, promote the free flow of information, and implement mechanisms of feedback and learning. See INST. MED., *supra* note 6, at 165-182.

A hospital's responsibility to the patient *vis-à-vis* supervision of staff, accreditation, and equipment has long been recognized in cases of corporate liability. In *Darling v. Charleston Community Memorial Hospital*,<sup>223</sup> the judicial decision often regarded as the origin of corporate liability, the Illinois Supreme Court was faced with the question of whether the hospital owed responsibilities directly to the patient. The Court ruled that such responsibilities were indeed desirable and feasible based on accreditation standards, state licensing requirements, and the hospital's own bylaws, as well as the expectations of the medical profession and other responsible authorities.<sup>224</sup> The court in *Darling* recognized for the first time that hospitals may incur liability for the negligent selection and monitoring of physicians who commit malpractice notwithstanding physicians' status as independent contractors, thereby establishing a legal duty on the part of hospitals to adequately credential physicians.<sup>225</sup> Along this line of reasoning, some have proposed that hospitals should be held liable for negligent credentialing when they fail to adequately vet clinical AI systems prior to clinical implementation.<sup>226</sup>

It should be noted that corporate liability, as distinct from strict liability coupled with CEL, is a fault-based regime that requires plaintiffs to prove hospital negligence as a prerequisite for recovery.<sup>227</sup> Such a fault-based approach poses significant obstacles to plaintiff recovery.<sup>228</sup> That being said, the case law on corporate negligence is instructive in affirming the role of the modern hospital as the patient's health care coordinator and, correspondingly, an important source of responsibility for harms that occur within its premises.<sup>229</sup> This generates responsibilities regarding the maintenance of equipment and establishment of adequate rules and policies.<sup>230</sup> In short, the idea that the hospital is particularly well-situated to minimize the risks that technology such as AI poses to patients is firmly established by the judicial recognition of corporate liability as a legal basis for holding hospitals liable to patients. This points toward the adoption of some form of EL for hospitals.<sup>231</sup>

#### D. STRICT VS. FAULT-BASED LIABILITY

The preceding discussion has suggested that fault-based liability regimes would place asymmetrical and unfairly onerous burdens on plaintiffs who seek to recover for harm arising from the use of clinical AI systems. While this unfairness and asymmetry favors a shift away from fault-based liability, it leaves open the question of what kind of no-fault or strict liability standard would be appropriate in these circumstances. There exist multiple instantiations of strict liability in American law, some more fault-like than others. On one end of the spectrum is modern products liability, which in some states resembles

<sup>223</sup>33 Ill.2d 326, 330, 211 N.E.2d 253, 256 (1965).

<sup>224</sup>*Id.* at 257, 332.

<sup>225</sup>Torin A. Dorros & T. Howard Stone, *Implications of Negligent Selection and Retention of Physicians in the Age of ERISA*, 21 AM. J.L. & MED. 383, 396 (1995). *Darling's* progeny has further defined the scope of a hospital duty to ensure patient safety and well-being. See, e.g., *Thompson v Nason Hosp.*, 591 A.2d 703, 707 (Pa. 1991), *aff'd by* *Brodowski v. Ryave*, 885 A.2d 1045 (Pa. 2005) (holding that a hospital has duties to: 1) to use reasonable care in the maintenance of safe and adequate facilities and equipment; 2) to select and retain only competent physicians; 3) to oversee all persons who practice medicine on hospital premises; and 4) to formulate, adopt, and enforce adequate rules and policies to ensure quality care for the patients).

<sup>226</sup>See Price II, *supra* note 85, at 304.

<sup>227</sup>Abraham & Weiler, *supra* note 162, at 391.

<sup>228</sup>*Id.*, at 391-92.

<sup>229</sup>See Mark E. Milsop, *Corporate Negligence: Defining the Duty Owed by Hospitals to Their Patients*, 30 DUQ. L. REV. 639, 660 (1991).

<sup>230</sup>See *id.* at 642-643.

<sup>231</sup>Abraham & Wiler, *supra* note 162, at 393.

negligence in requiring plaintiffs to demonstrate “fault-infused” elements such as reasonableness, foreseeability, and causation.<sup>232</sup> On the other end of the spectrum is a *Rylands v. Fletcher*-style<sup>233</sup> regime that imposes strict liability for harm resulting from “abnormally dangerous” activity.<sup>234</sup> Under this common law tort, liability applies even if the owner takes the appropriate precautions to prevent this risk from materializing. The *Rylands* rule was eventually incorporated into American law and today applies most commonly to damage resulting from activities such as the use of explosives and the transportation of nuclear materials.<sup>235</sup>

There are strong normative reasons to adopt some form of what Vladek calls “true strict liability,” which dispenses with legal tests found under products liability law and negligence.<sup>236</sup> Most crucially, providing victims redress for injuries sustained through no fault of their own, even if they cannot demonstrate elements such as foreseeability and causation, is consistent with “basic notions of fairness, compensatory justice, and the apportionment of risk in society.”<sup>237</sup> This applies especially in the medical space given patients’ lack of bargaining power in opting for the use of AI systems in the first place. Patients are rarely in a position to “negotiate all aspects of treatment, where, for example, they may consent to procedure without full comprehension of the procedure and its risks.”<sup>238</sup> Nor do patients have the ability to validate the soundness of clinical AI algorithms. On this point, it bears noting that strict liability has long been justified on normative grounds as a means of addressing the power imbalance between manufacturers and consumers.<sup>239</sup> In dispensing with the requirement that the tortfeasor and victim be linked together by the elements of fault or defect, strict liability also finds parallel in theories of distributed morality, espoused by scholars such as Luciano Floridi, where good or evil outcomes can be the “result of otherwise morally-neutral or at least morally negligible [] interactions among agents, constituting a multiagent-system, which might

---

<sup>232</sup>See Bryan Casey, *Robot Ipsa Loquitur*, 108 GEO. L.J. 225, 266 (2019) (“Products liability tests in force in the majority of states turn on principles of reasonableness, foreseeability, and causation that are congruent with findings of fault in negligence.”). This fault-infused liability standard leads Casey to call strict liability “zombie” regime that continues to cause analytic confusion.

<sup>233</sup>(1868) 3 LRE & 1 App. 330 (HL). *Rylands* held that a person can be liable for damage to a neighbor’s property flowing from the “non-natural” use of one’s own property. The idea is that while a property owner is free to store objects that have the propensity to escape and cause mischief to neighboring lands, this is done at the property owner’s own (legal) peril.

<sup>234</sup>See RESTATEMENT (SECOND) OF TORTS § 520 (AM. L. INST. 1977); RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL & EMOTIONAL HARM §20 (AM. L. INST. 2010).

<sup>235</sup>See John C. P. Goldberg & Benjamin C. Zipursky, *The Strict Liability in Fault and the Fault in Strict Liability*, 85 FORDHAM L. REV. 743, 761 (2016).

<sup>236</sup>See Vladek, *supra* note 5, at 146.

<sup>237</sup>*Id.*

<sup>238</sup>Kilcullen, *supra* note 170, at 15.

<sup>239</sup>See *Greenman v. Yuba Power Prod., Inc.*, 377 P.2d 897, 901 (Cal. 1963) (holding that the very purpose of products liability law is to countervail the power imbalance that exists between manufacturers and injured persons who are powerless to protect themselves). Despite the revolutionary potential of the *Greenman* decision, the promise of strict liability went unfulfilled due to doctrinal confusion over the meaning of “defect” and the tendency of courts to look for fault even when applying a strict-liability standard. See Andrzej Rapaczynski, *Driverless Cars and the Much Delayed Tort Law Revolution* 20 (Columbia L. & Econ. Working Paper No. 540, 2016), [https://scholarship.law.columbia.edu/faculty\\_scholarship/1962/](https://scholarship.law.columbia.edu/faculty_scholarship/1962/) [<https://perma.cc/NGQ5-CDSA>] (arguing that the best way to operationalize strict liability would have been to ask “who was most likely to be able to bring about safety improvements in the future, even if such improvements were not yet possible and even if we could not as yet specify them with any degree of precision. In other words, the relevant question is: Do we expect technical improvements in the design and/or the manufacturing process to be the best way of lowering the future cost of accidents of the type at issue, or do we expect some improvements to come from a more skillful or better calibrated use by the consumers, from medical advances in predicting or treating the injuries, or perhaps from some other inventions or behavior modifications?”).

be human, artificial, or hybrid.”<sup>240</sup> Under such a theory, responsibility for good and evil outcomes (the system’s outputs) can be backpropagated to all of the system’s nodes and agents for the purposes of improving the outcome.<sup>241</sup>

At the same time, it may not be appropriate to treat the use of clinical AI systems as a *Rylands*-style ultrahazardous or abnormally dangerous activity. The Third Restatement qualifies an activity as “abnormally dangerous” if “(1) the activity creates a foreseeable and highly significant risk of physical harm even when reasonable care is exercised by all actors; and (2) the activity is not one of common usage.”<sup>242</sup> While it is debatable whether the use of clinical AI system creates a “highly significant risk of physical harm” unsusceptible to mitigation through reasonable care, the criteria of non-common usage will be increasingly difficult to meet as this technology becomes more prevalent. If clinical AI systems advance to a stage where they match or even exceed physician performance, then it is possible that the combination of human and AI will become the standard of care.<sup>243</sup> To characterize medical AI as abnormally dangerous would at this point amount to saying that the practice of medicine is abnormally dangerous.

A shift away from fault-based liability would ultimately facilitate recovery from manufacturers and health care institutions for harmful results that are unpredictable, and to a large extent, unavoidable. The result would be to shift the cost of AI accidents to these enterprises. In the case of CEL, the costs would then be spread among the members of the enterprise, i.e., the actors most relevant to the creation, operation, and surveillance of the clinical AI system.<sup>244</sup> On this point, courts have long recognized the distributive goals that can be realized through a strict liability regime.<sup>245</sup> Moreover, a strict liability regime is more predictable and, as such, may be more conducive to innovation than a liability regime

---

<sup>240</sup>Luciano Floridi, *Distributed Morality in an Information Society*, 19 SCI. & ENG’G ETHICS 727, 729 (2012) [hereinafter *Distributed Morality*]. His only requirement for an agent to be included in such a system is that the agent exercise some degree of autonomy, interact with other agents and their environment, and is capable of learning from these interactions.

<sup>241</sup>Luciano Floridi, *Faultless Responsibility: on the Nature and Allocation of Moral Responsibility for Distributed Moral Actions*, 374 PHIL. TRANSACTIONS ROYAL SOC’Y A 2-3 (2016) [hereinafter *Faultless Responsibility*]. Floridi notes that reaching a satisfactory output in a social network is “achieved through hard and soft legislation, rules and codes of conducts, nudging, incentives and disincentives; in other words, through social pushes and pulls.” *Id.* at 7. This resonates strongly with the deterrence and distributive functions of tort law.

<sup>242</sup>RESTATEMENT (THIRD) OF TORTS: LIABILITY FOR PHYSICAL & EMOTIONAL HARM §20 (AM. L. INST. 2010). This formulation more or less encapsulates the six factors set out in the Second Restatement for what counts as an abnormal activity: “(a) Existence of a high degree of risk of some harm to the person, land or chattels of others; (b) Likelihood that the harm that results from it will be great; (c) Inability to eliminate the risk by the exercise of reasonable care; (d) Extent to which the activity is not a matter of common usage; (e) Inappropriateness of the activity to the place where it is carried on; and (f) Extent to which its value to the community is outweighed by its dangerous attributes.” RESTATEMENT (SECOND) OF TORTS § 520 (AM. L. INST. 1977). The term “abnormally dangerous activities” replaced the First Restatement’s language of “ultrahazardous activities”, though the substance of this latter category remains in the law. See Goldberg & Zipursky, *supra* note 235, at 760.

<sup>243</sup>We may even reach a point where every hospital and insurance company require the use of AI systems, with the failure to do so being legally actionable in the event of a bad outcome. See Froomkin et al., *supra* note 78, at 49-50.

<sup>244</sup>To borrow Floridi’s language, these are parties who “output, as a whole, a distributed action that is morally-loaded, by activating themselves and by interacting with other agents according to some specific inputs and thresholds, in ways that are assumed to be morally neutral.” *Faultless Responsibility*, *supra* note 241, at 7. CEL coupled with strict liability can be interpreted as a legal expression of this line of moral reasoning.

<sup>245</sup>See, e.g., *Escola v. Coca Cola Bottling Co. of Fresno*, 150 P.2d 436, 440, 443-44 (Cal. 1944) (Traynor, J., concurring) (proposing a shift from negligence to a strict liability standard for defective products based on the public policy that manufacturers are the best situated to anticipate product hazards. Moreover, manufacturing processes are often secretive and the consumer lacks the means to investigate a product’s soundness on their own); see also Vladeck, *supra* note 5, at 146 (characterizing strict liability as a “court-compelled insurance regime to address the inadequacy of tort law to resolve questions of liability that may push beyond the frontiers of science and technology”).

premised on the “quixotic search for, and then assignment, of fault.”<sup>246</sup> Like products liability, a CEL-based approach coupled with strict liability would recognize that decisions resulting in alleged harms can ultimately be traced upstream to choices made by manufacturing companies. By shifting at least some of the blame to the manufacturer without imposing on the plaintiff the burden of proving a defect or fault, the proposed approach would provide a strong incentive to take care in the design, programming, and manufacturing of clinical AI.<sup>247</sup>

The distributive objectives that strict liability seeks to further is not, contrary to what some commentators might think, entirely divorced from the corrective notion of justice privileged by fault-based torts such as negligence. Corrective justice seeks to “restore a pre-existing relationship between two parties, one that was unjustly disturbed by one party’s misconduct and the resulting injury to the other.”<sup>248</sup> It is commonly on this basis that commentators reject strict liability standard for AI harms; the critique is that it would be unfair to hold the defendant liable for harms over which they had no control and were, as such, unavoidable.<sup>249</sup> However, this is a misguided criticism, as it assumes that by holding defendants liable for unavoidable harms, it punishes a blameless party. Under the internal morality of harm-based strict liability, the wrong done is not the harm itself (which might very well have been unavoidable) but the failure to repair the harm done.<sup>250</sup> As Gregory Keating observes, “[t]he primary duty that harm-based strict liability institutes is *not* a duty not to harm; it is a duty to harm only through reasonable, justified conduct, and to make reparation for any harm done even though due care has been exercised.”<sup>251</sup> In this sense, strict liability does fulfill a corrective role. By holding the defendant liable for a failure to make reparation while benefiting from the activity that gave rise to the harm, strict liability restores the pre-existing relationship between the two parties.<sup>252</sup>

At the core of strict liability’s internal morality is a distributive idea of justice whereby it is wrong to inflict harm on another party—and benefit from that

<sup>246</sup>Vladeck, *supra* note 5, at 147.

<sup>247</sup>Liability could, for instance, incentivize manufacturers to make their code “crashworthy” in incorporating “state-of-the-art techniques in software fault tolerance.” See Bryan H. Choi, *Crashworthy Code*, 94 WASH. L. REV. 39, 47 (2019).

<sup>248</sup>AM. L. INST., *Perspectives on the Tort System and the Liability Crisis*, REPORTERS’ STUDY V: THE INSTITUTIONAL FRAMEWORK 3, 25 (1991).

<sup>249</sup>See e.g., Bathae, *supra* note 36, at 931.

<sup>250</sup>Under Keating’s taxonomy, harm-based strict liability – such as products liability law, abnormally dangerous activity law, and nuisance law – addresses *justifiable* conduct causing physical harm. In contrast, autonomy-based strict liability – such as trespass and battery – address *innocent* or morally blameless conduct that infringes on autonomy rights (over persons and property). In both instances, “[t]he object of the law’s criticism is not the defendant’s primary conduct in inflicting injury, but his secondary conflict in failing to repair the harm justifiably inflicted.” Gregory Keating, *Is There Really No Liability Without Fault? A Critique of Goldberg & Zipursky*, 85 FORDHAM L. REV. RES GESTAE 24, 30 (2016-2017).

<sup>251</sup>Gregory C. Keating, *Products Liability as Enterprise Liability*, 10 J. TORT L. 41, 66 (2017).

<sup>252</sup>*Cf.* Goldberg & Zipursky, *supra* note 235, at 766-767. Goldberg and Zipursky argue that Keating’s idea of “conditional wrongs” is untenable on the basis that a plaintiff does not have to prove that the strictly liable defendant failed to offer to pay for the damage caused. Their position is that the “predicate of liability is the doing of the harm, *not* the doing of the harm *plus* the failure to step forward to offer to pay.” Any pre-emptive payment from the defendant would be a matter of restitution, which presupposes the existence of a tort. Keating’s response is simple but, I think, effective: “[Golding and Zipursky] are right that no such proof is needed. Plaintiff need only prove that the defendant harmed her. The duty to repair the harm arises when harm is inflicted. If plaintiff and defendant cannot agree on what such reparation requires, the matter is for a court to determine simply because no one can unilaterally determine that they have discharged their legal obligations.” Gregory Keating, *Liability Without Regard to Fault: A Comment on Goldberg & Zipursky* 8 n.41, (Univ. of S. Cal. L. Sch., Working Paper No. 232, 2016) <https://law.bepress.com/cgi/viewcontent.cgi?article=1367&context=usclwps-iss> [<https://perma.cc/JQ5A-LMJ3>]

harm—without making reparations.<sup>253</sup> EL theories (including their CEL variants) embody most fully this principle of fairness in its focus on *activities* over individual actions. These are otherwise legitimate activities carried out by firms that, by their nature, expose consumers, employees, or the public to generalized and systemic risk. Activities in this sense are contrasted with the actions of separate, unrelated, and individual actors acting independently from each other that negligence law is most concerned with. The introduction of these kinds of risks does not constitute the activity itself but is, even for the most responsible enterprises, a necessary by-product that renders the activity possible and/or economically profitable. EL assumes that even if all reasonable precautions were followed, the enterprise bears a moral responsibility to make reparations for accidents due to the benefit accrued from the activity. It is this activity that serves as the basic unit of responsibility.<sup>254</sup>

Strict liability is fair to victims who, even if they are participants in the enterprise, do not benefit from the enterprise in proportion to the harm suffered as a result of the materialized risk. Strict liability is also fair to tortfeasors because it forces them to bear the cost of their choice to introduce these risks, which was presumably done in pursuit of their advantage.<sup>255</sup> This internal morality of EL is particularly germane to the medical context. If AI systems turn out to be more accurate than human physicians, their widespread use will likely drive down medical costs (including malpractice insurance) in the long term—a clear financial benefit for health care providers and organizations.<sup>256</sup> While patients will no doubt benefit from the use of health AI, this benefit is disproportionate to the detriment they suffer when physically harmed by an incorrect diagnosis or treatment recommendation.<sup>257</sup> The clear organizational and financial value derived by health care actors justifies a shift in who bears the cost of accidents.

#### E. FUTURE LEGAL REFORMS

Reforms to implement the proposed approach will likely have to take place at the state level given that tort actions, including those against hospitals, physicians, and manufacturers, have traditionally been the domain of state courts. Given the current stage of AI technology and lack of litigation involving clinical AI systems, it may be premature for state legislatures to mandate the replacement of product liability and medical malpractice law with CEL coupled with strict liability for harms arising out of the use of clinical AI systems. Instead, legislatures can pass statutes explicitly permitting hospitals and hospital systems to experiment with a CEL-based framework.<sup>258</sup> By comparing the effects of a hospital that assumes CEL for AI harms with a similar organization that does not, we can identify the effects of a CEL framework on metrics such as patient safety and number of lawsuits.<sup>259</sup> The results of these experiments can

---

<sup>253</sup>Keating, *supra* note 250, at 70.

<sup>254</sup>*Id.* at 72-74.

<sup>255</sup>*Id.* at 71.

<sup>256</sup>See Froomkin, et al., *supra* note 78, at 64. (“[W]e presume that [Machine Learning] diagnostics will follow the path of many other digital technologies and exhibit high fixed costs but relatively low marginal costs”).

<sup>257</sup>Keating, *supra* note 251, at 71.

<sup>258</sup>Proponents of EL have long advocated for these sorts of experiments. See, e.g., Paul C. Weiler, *Reforming Medical Malpractice in a Radically Moderate – and Ethical – Fashion*, 54 DEPAUL L. REV. 205, 231 (2005) (proposing that professional athletes’ associations, such as the National Hockey League Players’ Association, experiment with a EL-style, no-fault regime for their players).

<sup>259</sup>See, e.g., Kessler, *supra* note 164, at 102.



form the basis for future statutory reforms making CEL the exclusive remedy for harms arising out of the use of clinical AI systems.

Under the proposed approach, members of the common enterprise might want to set out the precise division of liability through the use of contractual indemnification clauses.<sup>260</sup> How liability is distributed in a given common enterprise will depend on a variety of factors such as the relative bargaining powers of the parties and the desire to promote innovation, though some liability on each party will be necessary to incentivize safety.<sup>261</sup> In practice, each actor's part of the claim will likely be paid by their respective insurers. Physicians and hospitals have access to coverage through the commercial insurance market, self-insurance, and the use of captive insurers.<sup>262</sup> While the kinds of insurance available to technology manufacturers typically exclude coverage for bodily injury, the insurance market is starting to close this gap.<sup>263</sup> Not only does the involvement of multiple insurers perform a useful loss-spreading function, but it may also promote patient safety in that insurers have a financial incentive to mandate AI safety requirements such as testing.<sup>264</sup> Admittedly, the involvement of a loss-absorbing entity in the form of the insurer appears to be in tension with the proposition that losses caused by an enterprise's activity ought to be borne by that enterprise.<sup>265</sup> This tension is mitigated, however, by the fact that medical liability insurance premiums and access to coverage have to a significant extent become risk-based. Liability insurers adjust premiums based on loss histories and may even refuse to insure high-risk medical providers.<sup>266</sup> As such, the introduction of liability insurance does not break the connection between liability and financial loss, even if it is the case that the members of the common enterprise would not bear the entirety of the loss.

At a certain point, the AI system may reach a level of autonomy and unpredictability such that we will have to reconsider the manufacturer's place in the common enterprise. While not essential to the proposed framework, recognizing a limited form of AI personhood would help deliver on some of the benefits of the CEL framework.<sup>267</sup> The AI "person" could be considered a participant in the common enterprise for the

---

<sup>260</sup> See Maliha et al., *supra* note 111.

<sup>261</sup> *Id.*

<sup>262</sup> See Hermer, *supra* note 171, at 297.

<sup>263</sup> Technology companies typically carry technology errors and omissions insurance, which is designed to cover financial loss and not bodily injury or property damage. General liability policy also excludes professional liability and therefore liability for bodily injury. Insurers have started to offer coverage for contingent bodily injury under technology errors and omissions policies, though at the moment only a limited number of insurers are willing to add this coverage. See Thompson Mackey, *Artificial Intelligence and Professional Liability*, RISK MGMT. MAG. (June 11, 2018), <http://www.rmmagazine.com/articles/article/2018/06/11/-Artificial-Intelligence-and-Professional-Liability> [<https://perma.cc/VER6-Z8H9>].

<sup>264</sup> George Maliha et al., *To Spur Growth in AI, We Need a New Approach to Legal Liability*, HARV. BUS. REV. (July 13, 2021), <https://hbr.org/2021/07/to-spur-growth-in-ai-we-need-a-new-approach-to-legal-liability> [<https://perma.cc/PH3H-YHGQ>].

<sup>265</sup> See, e.g., Helen Smith & Kit Fotheringham, *Artificial Intelligence in Clinical Decision-Making: Rethinking Liability*, 20 MED. L. INT'L 131, 148 (2020) ("... it is debatably not enterprise liability if the economic impact of a claim is assigned to an insurer rather than directly impacting the actors who caused the harm").

<sup>266</sup> See Ted Baker & Charles Silver, *How Liability Insurers Protect Patients and Improve Safety*, 68 DEPAUL L. REV. 209, 237 (2019).

<sup>267</sup> See Vladeck, *supra* note 5 ("Conferring 'personhood' on those machines would resolve the agency question; the machines would become principals in their own right and along with new legal status would come new legal burdens, including the burden of self-insurance. This is a different form of cost-spreading than focusing on the vehicle's creators, and it may have the virtue of necessitating that a broader audience – including the vehicle's owner – participate in funding the insurance pool, and that too may be more fair").

purposes of liability—taking the place of the manufacturer and component suppliers. Gerhard Wagner has noted that robot personhood can serve to “bundle” responsibility and allow liability to be attributed to a single entity to which the victim may turn for compensation.<sup>268</sup> Like corporate personality, AI personality is one way of ensuring accountability where the harm can be traced to the activities of a group but not to any single individual. It would also avoid the complication of having one common enterprise (among the AI manufacturer and subcomponent manufacturers as envisioned by Vladeck) form part of another larger common enterprise (among the physician, manufacturer, and the hospital).

Many of the concerns about AI personhood can be allayed either with mandatory liability insurance and/or assets backing the AI person.<sup>269</sup> Karnow has proposed the creation of a Turing Registry that would certify the risk level of an AI system, charge the developer a commensurate premium for liability coverage, and pay compensation for harms without any inquiry into fault or causation.<sup>270</sup> Similarly, the EU parliament has recommended the creation of an obligatory AI insurance, along with a supplemental insurance fund, as corollary to their call for AI personhood.<sup>271</sup> A more limited version of this proposal could be adapted for health AI systems with members of the common enterprise as the payors. The AI person would therefore function as little more than a conduit to directly channel the costs of insurance to certain actors.<sup>272</sup>

## V. CONCLUSION

While WFO’s recent misadventures should not scare us off from exploiting the clinical promise of AI technology, the legal system would be wise to prepare for novel legal disputes involving AI-related harms in the health care space. Due to their complexity, opacity, and lack of foreseeability, AI systems are not easily accommodated by traditional liability frameworks. As such, frameworks based on fault or defects will make it difficult for victims of AI harms to obtain compensation. Agency law, on the other hand, is too limited of a legal theory to account for the dispersion of responsibility that characterizes the operation of clinical AI systems. A key insight of this Article is that clinical AI is deeply intertwined with the operation, mission, and expertise of health care organizations. Given this background, applying CEL would be both fitting and desirable. By recognizing

---

<sup>268</sup>Gerhard Wagner, *Robot, Inc.: Personhood for Autonomous Systems?*, 88 *FORDHAM L. REV.* 591, 608 (2019).

<sup>269</sup>*See id.*, at 610; EUR. COMM’N, *supra* note 4, at 38 (holding that “[a]ny additional personality should go hand-in-hand with funds assigned to such electronic persons, so that claims can be effectively brought against them. This would amount to putting a cap on liability and – as experience with corporations has shown – subsequent attempts to circumvent such restrictions by pursuing claims against natural or legal persons to whom electronic persons can be attributed, effectively ‘piercing the electronic veil’”).

<sup>270</sup>Karnow, *supra* note 189, at 193-196.

<sup>271</sup>Comm. on Legal Affairs, *supra* note 140, at ¶¶ 58-59. Unlike Karnow’s proposal, the EU’s insurance scheme does not include risk certification.

<sup>272</sup>*See* Wagner, *supra* note 268, at 610 (“If the manufacturers have to front the costs of insurance, they will pass these costs on to the buyers or operators of the robot. In one form or another, they would end up with the users. The same outcome occurs if users contribute directly to the asset cushion or become liable for insurance premiums. In the end, therefore, the robot’s producers and users must pay for the harm the robot cases. The ePerson is only a conduit to channel the costs of coverage to the manufacturers and users.”); *see also* Karnow, *The Application of Traditional Tort Theory to Embodied Machine Intelligence*, *supra* note 77, at 51 (“In an age of mass markets and long distribution chains, costs could be allocated across a large number of sales, and manufacturers were in a position accordingly to spread costs including by purchasing insurance. Why not similarly spread the costs of injury?”).

a common enterprise among physicians, AI manufacturers, and hospitals, the law can address the threat of a responsibility gap and leverage the hospital's unique influence over the safe use of health technology. The proposed framework's shift away from fault-based liability serves a deterrent function while favoring victim compensation. Notwithstanding unresolved issues, including whether AI personhood could help deliver on some of the benefits of the proposed approach, a move towards CEL would likely facilitate the adoption of clinical AI technology while ensuring fair compensation for harms arising out of the technology's use.