Locating Liability for Medical AI

W. Nicholson Price II
I. Glenn Cohen

Follow this and additional works at: https://via.library.depaul.edu/law-review

Part of the Law Commons

Recommended Citation
W. N. Price II & I. G. Cohen, Locating Liability for Medical AI, 73 DePaul L. Rev. 339 (2024)
Available at: https://via.library.depaul.edu/law-review/vol73/iss2/8

This Article is brought to you for free and open access by the College of Law at Digital Commons@DePaul. It has been accepted for inclusion in DePaul Law Review by an authorized editor of Digital Commons@DePaul. For more information, please contact digitalservices@depaul.edu.
LOCATING LIABILITY FOR MEDICAL AI

W. Nicholson Price II & I. Glenn Cohen*

When medical AI systems fail, who should be responsible, and how? We argue that various features of medical AI complicate the application of existing tort doctrines and render them ineffective at creating incentives for the safe and effective use of medical AI. In addition to complexity and opacity, the problem of contextual bias, where medical AI systems vary substantially in performance from place to place, hampers traditional doctrines. We suggest instead the application of enterprise liability to hospitals—making them broadly liable for negligent injuries occurring within the hospital system—with an important caveat: hospitals must have access to the information needed for adaptation and monitoring. If that information is unavailable, we suggest that liability should shift from hospitals to the developers keeping information secret.

Table of Contents
INTRODUCTION .................................. 340
I. MEDICAL AI, INJURY, CONTEXTUAL BIAS, AND
HUMAN OVERSIGHT ............................... 343
A. Failure and Contextual Bias ...................... 345
B. Human Oversight ................................ 347
II. WHY TORT CLASSIC® ISN’T HELPFUL .......... 348
A. Physician Liability ............................... 349
B. Developer Liability ............................... 350

* Nicholson Price is a Professor of Law at the University of Michigan. Glenn Cohen is the James A. Attwood and Leslie Williams Professor of Law; Deputy Dean; and Faculty Director, Petrie-Flom Center for Health Law Policy, Biotechnology & Bioethics, all at Harvard Law School. We thank Kenneth Abraham, Oren Bracha, Ana Bracic, Bryan Choi, Charlotte Alexander, Sherman Clark, Rebecca Crootof, Deven Desai, Greg Day, Rebecca Eisenberg, John Goldberg, Yuan Hao, Scott Hershovitz, Rachel Rothschild, Peter Salib, Matt Sag, Catherine Sharkey, Ben Sundholm, Charlotte Tschider, and participants at the NYU Artificial Intelligence: Tort and Administrative Law Seminar, the Michigan Law School Fawley Workshop, the Emory AI and Law Roundtable, and the Clifford Symposium at DePaul University College of Law for useful comments. We thank Pallavi Bugga for excellent research assistance. This work was funded in part by a Novo Nordisk Foundation Grant for a scientifically independent International Collaborative Bioscience Innovation & Law Programme (Inter-CeBIL programme - grant no. NNF23SA0087056). Cohen also discloses that he is a member of the ethics advisory board for Illumina and the Bayer Bioethics Council. He was also compensated for speaking at events organized by Philips with the Washington Post and retained as an expert witness in health privacy lawsuits.
AI systems fail. When they do, sometimes people get hurt. How should the tort system deal with these failures to compensate the injured and minimize injuries in the first place? We’re not the first to consider these issues (not by a long shot), but we aim to highlight the complexities of AI and tort law in the particular context of medical injuries, and against a particularly important background point: The performance of AI systems (and here we’re talking mostly about machine learning, an important and fast-moving subset of AI) varies substantially based on the characteristics of the data on which the system was trained, the characteristics of the place the AI system is used, and the process of translating from the first to the second. These performance differences, we argue, make a substantial difference to how tort law should treat failures of medical AI.

When new technological challenges are claimed to create issues for tort law, a common answer is: Why the fuss? Tort law is flexible, has dealt with lots of things in the past, and will deal with this too. And indeed, it’s tempting to rely on well-established common-law principles of tort to allocate responsibility, compensate wrong, and create incentives to avoid injury here, adjusting flexibly to new circumstances as they have to so many innovations in the past. Under such an approach, physicians would be held liable if they fell below the standard of care in their treatment, regardless of what the AI system said, unless and until AI use becomes the standard of care; hospitals might be liable vicariously for physician actions or directly for selecting AI systems poorly; and developers might—might—be liable in product liability for

---

negligently designing or manufacturing AI systems or failing to warn of their limitations.\(^2\)

Unfortunately, traditional tort liability has real problems setting the right incentives to limit injuries with medical AI, especially with issues that arise from differential performance. Medical AI is opaque, changeable, and brittle—it’s hard to see what is happening, what is happening can change over time, and performance can readily drop in different circumstances. This combination is remarkably ill-suited for physicians, the most obvious site of liability, to address. They do not, and reasonably will not, have the time, expertise, or opportunity to catch problems with medical AI.\(^3\) And if medical AI is meant to allow health systems to do more, or perform better, than their staff would otherwise allow, it’s particularly pernicious to imagine that overworked and undertrained providers could catch these kinds of errors. Allowing liability for such systemic errors fall on providers seems both unjust and unhelpful.

As between hospitals and developers, the problem is one of mutually insufficient information. Health systems know their staff, resources, and patient populations (or they should, and should be held to)—but they don’t know how AI systems are trained, and frequently the details of training and datasets will be actively kept secret by developers.\(^4\) Developers, on the other hand, know how their AI systems work, but know little about the nitty-gritty details of various health system issues that turn out to matter a great deal to how the AI system will perform in real life. Neither party is precisely in position to be the cheapest cost avoider.

We propose to cut this Gordian knot by resting liability for malfunctions on medical AI squarely on the shoulders of the health systems that deploy them, with one caveat that we’ll mention in a moment. This allocation of responsibility has three main justifications. First, the health system has the last-mile knowledge that developers don’t, and ultimately the last mile is where the patients are. Second, allocating responsibility to health systems mostly deals with the lingering question of physician malpractice—an increasing number of physicians are now employed rather than independent contractors, and the use of centrally managed AI systems to guide care is perhaps the epitome of care coordination.

---

2. See infra Part II.B.; see also W. Nicholson Price II, Sara Gerke & I. Glenn Cohen, Liability for Use of Artificial Intelligence in Medicine, in RESEARCH HANDBOOK ON HEALTH, AI AND THE LAW (Barry Solaiman & I. Glenn Cohen eds.) (forthcoming 2024) [hereinafter Price II et al., Liability for Use of AI] (summarizing these theories of liability). While we refer to “hospitals” in general, for liability purposes in some instances their components may need to be disaggregated such as in the case of freestanding emergency rooms which might be owned and operated by other parties.


at the heart of direct hospital liability. Third, leaving medical AI liability with the health system avoids the difficult question of disentangling problems that arise from original training, system maintenance and drift over time, and problems integrating the system into the clinical workflow. More specifically, we propose adopting the enterprise liability models others have urged should be imposed on hospitals for injuries sustained in hospitals in general—models that would make hospitals broadly liable for negligent injuries occurring within the hospital system—to the limited case of injuries sustained in hospitals from medical AI.

Here's the caveat: for liability to be imposed, the health system needs to be able to interrogate the model and the data on which it was trained. If developers conceal their training parameters and datasets, so that hospitals cannot reasonably evaluate and tune the models, liability should instead shift to developers. This caveat has the salutary effect of promoting disclosure, which should help with effective governance of AI systems more generally.

This solution is not perfect, of course. Among other things, health systems may lack the resources to effectively check local performance (a problem worth tackling separately). But crisply allocating liability brings an additional benefit: parties can readily bargain to shift liability if confident that they can cheaply ensure good performance in the relevant context. And in fact, some parties already do this; Digital Diagnostics, for instance, contractually indemnifies providers that use its autonomous IDx-DR diabetic retinopathy detection system (and carries its own malpractice insurance). To be sure, as we discuss at the end of this Article, there are many questions left open, including how to deal with cases where mix of AI and non-AI errors cause injury.

5. We note that enterprise liability could also be strict, rather than negligence-based. While this is not the primary focus of our Article, and while we do not take a definitive position, there are some strong arguments others have offered as to why should be skeptical of strict liability approaches in the medical injury setting. See, e.g., Richard A. Epstein, Medical Malpractice: The Case for Contract, 1 Am. Bar Found. Rsch. J. 87, 102–05 (1976). There are some important differences between this setting and those where strict liability is championed, such as auto accidents or workers’ compensation regimes. Defining the compensable event, such as an injury arising out of medical treatment, is difficult in the medical context because people who get treatment are already sick or injured—unlike the average driver who enters her car or the average worker who enters his workplace uninjured. Relatedly the costs to determine causation tend to be higher than in other areas. We thank Benjamin Zipursky and Kenneth Abraham for discussions on this point.

6. See W. Nicholson Price II, Distributed Governance of Medical AI, 25 SMU Sci. & Tech. L. Rev. 3, 16–17 (2022) [hereinafter Price II, Distributed Governance of Medical AI] (describing different capacity); Id. at 18–21 (proposing mechanisms to improve low-resource health-system capacity).

One brief note on scope. Reducing the rate of injuries can be accomplished via multiple levers; one is ex post tort liability, but another is ex ante regulation (and of course there are more). We focus here on tort liability, but recognize that this is only part of the picture, and not an isolated part at that. (For what it’s worth, ex ante regulation faces real difficulties that are the subject of other work.)

I. Medical AI, Injury, Contextual Bias, and Human Oversight

We begin by taking a step back to briefly canvas what AI is doing in medicine and how it can harm patients. Medical AI has moved remarkably rapidly, from a blue-sky pipe dream to products that are actually in use and affecting care for many millions of patients in various ways. AI systems are already involved in predicting patient deterioration, helping optimize health system staffing and patient flow, analyzing vital signs (and other signals like video or audio) for likely signs of illness, helping to make (or autonomously making) diagnoses based on image analysis, and recommending treatments. Some AI systems are developed in-house by health systems, many are created by providers of...
electronic health record (EHR) software, and many are made by commercial developers.\textsuperscript{10} Hundreds of AI systems have been cleared for use by the U.S. Food and Drug Administration (FDA), though many systems in use have not gone through any FDA oversight.\textsuperscript{11} These systems are quickly entering patient care in health systems across the country (and the world).

But AI systems aren’t perfect, and when they go wrong, patients can be harmed. If an AI system misses a diagnosis, the patient may not seek the proper care. If the system recommends an incorrect drug, the patient may forego benefits, suffer unnecessary side effects, or suffer more serious harm. If the system predicts risk inaccurately, staff might not watch the patient carefully enough—or might watch too closely, shifting time and effort from other patients who might need it more. At least some of these harms are cognizable as injuries within the tort system.\textsuperscript{12} There is copious evidence of medical AI systems making mistakes. Epic, the largest vendor of electronic health record systems, touted an embedded predictor of sepsis risk that performed poorly in many places;\textsuperscript{13} most of the predictors trying to guide care during the beginning of the COVID-19 pandemic failed to be of any use;\textsuperscript{14} and a system used to predict care needs for many millions of patients was deeply racially biased.\textsuperscript{15}

For now, harms from AI are more theoretical than demonstrated, for two main reasons. First, timing. While AI systems are entering care workflows, they’re far from omnipresent, so there are simply fewer patients to be injured and fewer ways to injure them—so far. Also, to the extent that it takes some time for injuries to occur, be traced to AI failures, and work their way through the system, that process delays the observation of AI-related injuries. Second, opacity. Medical errors are already difficult to observe against a background of already sick, hurt people who are treated with care that often doesn’t work, even when performed

\textsuperscript{10} Price II et al., \textit{New Innovation Models in Medical AI, supra} note 8, at 1125–26.
\textsuperscript{11} \textit{Id.} at 1126–28.
\textsuperscript{12} Questions of scarce resource allocation between different patients are a rather awkward fit for the tort system with its emphasis on individuals harmed by individual actions; we set these troublesome, if fascinating, issues aside because there are plenty of harms that are much more straightforwardly injuries in the tort sense.
perfectly. The vast majority of malpractice goes unremarked. Further, when litigation is brought it often settles and, as others have observed, the cases that litigate to court determination are quite an imperfect guide to whether liability should lie. The opacity of medical AI, where even its creators may not know precisely how it makes recommendations or predictions, makes it even harder to link an individual patient injury to an underlying failure of AI. There are certainly harms that have arisen from the use of AI systems in clinical care—a prominent paper discusses a racial bias that led some patients not to receive beneficial care coordination while others did—but the known harms remain limited to statistical, rather than identified, individuals.

A. Failure and Contextual Bias

So why does medical AI fail? Alongside a long list of standard sources of error (programmers make mistakes, data aren’t adequately cleaned, system outputs are unclear or are incorrectly interpreted), a doggedly persistent problem arises from the challenge of generalizability and contextual bias: AI systems developed in one place often perform differently, and often perform worse, when they are used in other places. Contextual bias is, in some senses, a flipside of a perceived value of medical AI: its consistency in technical performance. Given the same inputs, AI should give the same outputs, without the inconsistency often introduced by variable human actors. That consistency can be

---


18. Obermeyer et al., supra note 15. Non-AI-algorithms used in care have resulted in similar statistical harms. See, e.g., Leila R. Zelnick, Nicolae Leca, Bessie Young & Nisha Bansal, Association of the Estimated Glomerular Filtration Rate With vs Without a Coefficient for Race With Time to Eligibility for Kidney Transplant, 4 JAMA Network Open e2034004 (2021) (noting that racial adjustments to kidney function formulae have resulted in patients being delayed on transplant lists).


20. We elide many technical questions about precision versus recall, Type I and Type II errors and how they might be treated different, the usefulness of ROC and AUC; we think the intuitions we capture here do not depend on these details, but welcome further exploration along those lines.

good—but not when it fails to pick contextual differences that matter. This problem is pervasive because of the way AI systems are created.

Medical AI systems are typically developed in relatively high-resource environments (wealthy hospital systems like Mass General Brigham, Mount Sinai, or Stanford) on data that reflect a small slice of the health system, because those are the data that are easiest to get. The vast majority of data used in medical AI, for instance, come from just three states (Massachusetts, New York, and California), and hundreds of AI papers use data from precisely one hospital, Beth Israel Deaconess in Boston—which makes available a high-quality, open-access dataset for research.

Development on concentrated data wouldn’t be a problem if AI systems performed just as well in places different from where they were developed, but that’s rarely the case. AI systems learn from the data they have, and health systems differ with respect to patient populations, care patterns, medical equipment, the training and availability of medical providers, and even the way information is recorded. When AI systems encounter different patients, providers, and data systems, they may perform exactly as intended—but they may also perform differently, perform worse, or break entirely. And patients may be injured as a result; for instance, if an AI system trained to predict sepsis in one hospital does poorly in another, that poor performance might lead to missed interventions and increased sepsis (especially if medical staff assume the system makes reliable predictions).


25. Price II, Distributed Governance of Medical AI, supra note 6, at 11–12.

26. Making matters worse, it can be difficult to predict in advance which will happen, and without careful monitoring (the type of monitoring that is frequently missing in health systems, especially those with fewer resources) it can be difficult even to measure how the AI system actually performs.

27. See Wong et al., supra note 13. A particularly knotty question arises: How should we think about cases where adopting an AI produces an overall benefit across a large population of patients but because of contextual bias in a particular patient’s case the result is worse for a particular patient? An all-things-considered perspective might care not just about the magnitude of the improvement in overall performance but the distribution of cases where the AI leads to worse performance than care without medical AI. It may matter, especially from the perspective of tort law, whether “worse” is merely worse than the AI system performs otherwise but still an improvement over the non-AI standard of care (maybe non-actionable) or worse than that non-AI standard of care (maybe actionable). The ethics of this result are complicated, and beyond our scope here. For an intriguing possibility of ethical analysis along these lines, see Peter N. Salib, Abolition by Algorithm (unpublished manuscript) (on file with authors).
B. Human Oversight

One last piece of the background puzzle: human oversight of AI systems, both now and in the future. For now, at least, AI systems are typically kept on quite a short leash, frequently limited to giving advice to health-care providers who are very much still the final decisionmakers. In fact, regulation and existing tort doctrine create incentives for keeping providers centrally in the loop, and for those providers to use AI systems mostly to confirm their existing practices rather than to do things they would not otherwise do.28

But there is little reason to expect that close human oversight of AI systems will continue—or to hope that it should. Humans working with machines frequently learn to defer to the machines over time; what seems like new technology today, with a need for careful oversight, is old hat in a few years, and oversight is likely to slip.29 Providers are strapped for time, and when machines make recommendations that take a long time to validate, that validation may frequently be pushed to the bottom of the pile.

Opacity is also a major hindrance to physicians overseeing AI. AI systems typically generate results using processes that are opaque or incomprehensible, even to those with significant technical expertise or their own developers. Developers can trade transparency for accuracy, but that’s a fraught trade.30 “Explainability” methods that purport to provide mostly-correct explanations for too-complicated-to-understand underlying processes risk unhelpfully oversimplifying,31 and some work has suggested that mostly-correct explanations are worse for physician performance than no explanations at all.32

More fundamentally, AI systems aren’t suited to one-off human oversight in two significant ways. First, they have the potential to perform beyond the limits of human performance (for instance, predicting

---

30. Joshua Hatherley, Robert Sparrow & Mark Howard, The Virtues of Interpretable Medical Artificial Intelligence, Cambridge Q. Healthcare Ethics 1, 1 (2022) (noting that if physicians and patients only follow changes if they understand them, the tradeoff between accuracy and transparency might not be so sharp).
acute kidney injury many hours before the best human possibly could, or cancer months before it is observable by oncologists on an MRI). Humans trying to correct individual decisions of an opaque system that performs better, on average, than any human is a challenging exercise, to say the least. Second, AI systems have the potential to democratize expertise, allowing practitioners to provide care beyond their level of expertise. The IDx-DR system, for instance, lets general practitioners diagnose more-than-mild diabetic retinopathy, something that could otherwise require an ophthalmologist. Correcting errors of something outside a provider’s expertise is, almost by definition, asking too much of the provider.

* * *

Putting this all together, AI is increasingly used in medical care, and while there hasn’t been much in the way of documented injury, failures and resulting injuries are inevitable. Systemically, failures are mediated by the fact of contextual bias, where AI systems perform differently—typically worse—outside the environments where they were developed and optimized. And individual human intervention at the moment of decision is a tenuous lever to minimize those errors. On to torts!

II. Why Tort Classic® Isn’t Helpful

Tort law’s default response to new technology is to yawn. Standards are flexible; reasonableness changes with time and with context; and existing doctrines cover liability for health-care injuries, even in situations with complex relationships between actors. We argue that, nevertheless, existing doctrine will have a remarkably tough time tackling medical AI, at least if we assume a goal of setting incentives right to lower the rate of injuries and improve patient safety.33 (Really! This time it’s different!) Existing doctrines are a poor fit for physician malpractice; hospital liability, whether vicarious or direct; and manufacturer liability. As Barbara Evans and Frank Pasquale put it, “When diagnostic AI ignores problems with inclusivity and bias yet still manages to deliver better results than unaided human observation for many or most patients, the patients who do suffer an injury may not have a tort remedy under a negligence standard . . .”34

33. Our focus is on the goal of reducing patient harm by reducing the incidence of medical error. That is not, of course, the only perspective and we leave open questions about how frameworks more focused on corrective justice or civil recourse might treat the problem.

(One quick aside: in the discussion that follows, we assume that manufacturers and hospitals are different entities; we’ll discuss later what happens if we relax that assumption and hospitals develop their own AI systems.)

A. Physician Liability

Physician\textsuperscript{35} liability, the most obvious place to tackle medical injuries, won’t do much to drive improvements in the safety and functioning of medical AI in practice. Physicians are, of course, liable for malpractice when their actions fall below the standard of care of a reasonable physician, typically determined according to the custom of similarly situated physicians.\textsuperscript{36} This applies to the use of AI as it does to the use of other medical technologies. We don’t know yet, precisely, what it means to be reasonable in using AI technology in patient care. In prior writing we’ve suggested that the presence or absence of an AI recommendation probably doesn’t change what treatment the standard of care demands, at least for now,\textsuperscript{37} suggesting that liability concerns will incentivize physicians to use AI largely to confirm doing what they already intend to do, and maybe to streamline or triage, rather than to try new things.\textsuperscript{38} We’ve also suggested that physicians might have a duty (or should have a duty) to know some reasonable things about how AI systems were developed, trained, and validated.\textsuperscript{39} All well and good, and hopefully enough to make physicians more thoughtful and careful about how they choose and use AI systems.

But realistically, physicians have little chance of fixing or compensating for AI problems of the type we have described, for all the reasons we have described above. Performance differences themselves are a huge barrier: if AI systems perform better than physicians on a specific task on average, and physicians follow their own judgment when they disagree with the AI, that result is (by construction!) worse just blindly following the AI system. Consider an unavoidably stylized example: a system recommends a diagnosis to a physician and gets it right 98%

\textsuperscript{35} We discuss physicians, but similar arguments apply to any provider. Physician malpractice has just been the best hashed out.

\textsuperscript{36} There is variation in how states articulate this standard, but we do not see that variation as particularly pertinent here. Since current standards take some account of resources—physicians in hospitals without MRI machines are less likely liable for failing to order an MRI—one of those standards might include something about the resources available for medical AI. But that’s deeper in the weeds than we consider in our current stroll through the doctrinal garden.

\textsuperscript{37} Some have suggested this will change fairly soon. See Froomkin et al., supra note 29, at 35–36.

\textsuperscript{38} Price II et al., Liability for Physicians, supra note 1, at 1765.

\textsuperscript{39} Price II et al., Liability for Use of AI, supra note 2.
of the time, when competent human physicians get it right 95% of the
time. If the physician follows the recommendations, the performance
is 98%. If the physician follows the recommendations when they agree
with them, and diverges when it disagrees with them, performance
drops back to 95%.\footnote{As a first approximation. It is more complicated, because physicians likely treat false nega-
tives differently than false positives, and the physician is likely to make at least some different
events than the AI system will.}

In such a circumstance, tort law gets a bit turned around. On the one
hand, it would be hard to find malpractice for following an AI system
that, on average, does better than a reasonable physician. On the other
hand, a reasonable physician would implement the standard of care
used by competent physicians—that is, the care that results in that 95%
success rate. Both seem reasonable, at least at first glance. If the answer
is that neither path leads to liability, that’s nice for physicians, but not
great at catching errors that could have been avoided. But posit contextual bias, and consider the hypothetical where the system showed 98%
performance in testing, but dropped (without anyone noticing) to 85%
in the setting where it’s being used. If tort law leads to physicians—reason-
ably?—deferring to the system without any idea of contextual bias,
then that’s a substantial problem.

Ultimately, liability for human physicians for malpractice certainly
influences behavior and performance around the edges, but doesn’t
have much purchase at fixing systemic problems for how AI is devel-
oped and deployed.

\section{Developer Liability}

If physician liability is unlikely to solve the problems we have set out,
one natural inclination is to go to the source and consider imposing
liability on developers instead. While developer liability is also possible
under existing tort law, it also fails to match capabilities to responsi-
bility for the purposes of decreasing injury—just on the opposite end
of the timeline from development to final use. We do not argue that
developer liability is totally unhelpful, but we view it as an incomplete
solution at best.

Tort law naturally turns to products liability, whether design defect,
manufacturing defect, or failure to warn. Under these theories, develop-
ers could be responsible for defects that occur in designing and training
the AI system (though frankly, it’s not obvious whether the latter more
naturally fits as design defect or a manufacturing defect), or for failure
to warn users about the limitations of the AI systems they develop.
There are a few problems with relying on this approach, too. First, courts have been consistently reluctant to treat software as a product, often judging it as a service instead.\textsuperscript{41} Evans and Pasquale suggest that FDA clearance of AI systems may ease the road for courts to treat these systems as products, though this path remains unexplored. This path is also limited on both ends: many AI systems do not go through significant FDA scrutiny\textsuperscript{42} (though this may change),\textsuperscript{43} and those systems that go through full Class III premarket approval processes (a tiny minority, at least for now) are likely have design defect suits preempted.\textsuperscript{44}

Second, and more significantly, developers have an information problem as well. Put simply, they lack information about the workings of the hospital where their systems will be deployed; they do not know the details of its patient populations, how hospital workflows function in practice, or how data structures there may change over time. To be sure, developers of some AI systems needn’t worry about these products; if a system uses the same camera every time to take pictures of patients’ eyes, no matter the hospital or patient, perhaps adequate training can virtually guarantee that performance will not change. But this will not always be possible.

\textsuperscript{\* \* \*}

Taking these factors together, existing tort doctrine provides a messy, ill-structured tangle of doctrines which are unlikely to effectively drive the application of safe, effective AI systems in health. In addition to informational mismatches, current doctrines are fogged by uncertainty of application, obscuring an already opaque landscape. Effective incentives through liability demand more of the tort system.

\section{Hospital Liability, Enterprise or Otherwise}

Given the issues with physician and developer liability, one enticing solution would be to move to hospital liability. We discuss existing theories of hospital liability, turn to enterprise liability as applied to hospitals, and then examine why it fits particularly well in the case of medical AI.

\begin{footnotesize}
\begin{itemize}
\item[41.] Evans \& Pasquale, \textit{Product Liability Suits for FDA-Regulated AI/ML Software, in The Future of Medical Device Regulation: Innovation and Protection, supra note 34}, at 24. Treating AI systems as a service raise the mirror image of problems with individual physician use: developers have too little information about individual instances to oversee them adequately.
\item[42.] Price II et al., \textit{New Innovation Models in Medical AI, supra note 8}, at 1141.
\item[43.] See generally Ariel Dora Stern, \textit{The Regulation of Medical AI: Policy Approaches, Data, and Innovation Incentives, in The Economics of Artificial Intelligence: Health Care Challenges} (Ajay Agrawal, Joshua Gans, Avi Goldfarb \& Catherine Tucker eds.) (forthcoming 2024).
\end{itemize}
\end{footnotesize}
A. Direct and Vicarious Liability

Before diving into the doctrine, it's worth taking a moment to note why hospital liability is unusual and more complex than other forms of service delivery. While businesses in most industries are typically responsible for the carelessness of their employees, when it comes to hospitals, physicians are at baseline “treated as independent contractors, rather than agents or employees of the hospital” such that “hospitals are not ordinarily vicariously liable for the negligence of treating physicians.”

Nevertheless, several lines of doctrine have created exceptions to this background rule. We briefly discuss them and mention how they may in the future be used as to medical AI, largely to contrast them with what adopting enterprise liability for medical AI would mean.

First, there are derivative liability theories of tort against hospitals based on establishing the medical malpractice or liability of the hospital’s physicians or other health care workers. Most obviously, an increasing number of physicians are now employed by hospitals; respondeat superior makes hospitals liable for their negligence.

However, many physicians remain independent contractors, not employees, such that respondeat superior is not available. Some courts have found ways around this in finding implicit control relationships justifying respondeat superior in the absence of formal employment, or in finding non-delegable duties for essential hospital functions, though establishing this may be challenging for plaintiffs. A different theory of “apparent authority” applies “when a third party reasonably believes the actor has authority to act on behalf of the principal and that belief is traceable to the principal’s manifestations.”

Through any of these vicarious liability theories, a patient could seek to prove medical malpractice of the physician for his or her use of medical AI and then try to bootstrap to

45. Philip G. Peters, Jr., Resuscitating Hospital Enterprise Liability, 73 Mo. L. Rev. 369, 373 (2008). Physicians largely favored this arrangement because it provided more independence from hospital oversight. Id. In fact, when the famous 1965 Illinois Supreme Court case of Darling v. Charleston Community Memorial Hospital threatened to change that status quo, 211 N.E.2d 253 (Ill. 1965), physicians cried foul, other courts did not follow, even Illinois’ courts pulled back, and in the years since organized medicine has successfully stemmed a wave of major change towards hospital liability. Id. at 373–76.


47. RESTATEMENT (THIRD) OF AGENCY § 2.03 (AM. L. INST. 2006); Popovich v. Allina Health Sys., 946 N.W.2d 885, 891 (Minn. 2020).
liability against the hospital. But, if as we have argued above, physician negligence law matches very poorly with the task of trying to ensure the performance of AI systems across different health-care environments, then there is no reason to believe bootstrapping from it up to hospitals will get things right.

Second, there are a series of direct hospital liability theories that might be tried by plaintiffs claiming injury from medical AI, most prominently negligent selection or retention and negligent supervision. In the first, a plaintiff succeeds by showing that a hospital did not exercise reasonable care in determining whether a physician or other health care worker was competent.\textsuperscript{48} As applied to AI, this would play on the metaphor that a hospital system is “hiring,” not “buying,” an AI system and that the hospital failed in its duties to adequately review the AI’s performance, as well as its integration into the hospital workflow.\textsuperscript{49} To date, we are aware of no case that has litigated such a claim. In negligent supervision, the breach is of a duty regarding “contemporaneous supervision of daily treatment decisions as they are made.”\textsuperscript{50} As applied to AI, such a theory might run into particular difficulties when the algorithm is on the opaque side such that far from “supervising” the decision, a physician may not even understand the basis for the decision. In any event, this theory of liability has been disfavored by the courts, typically mentioned in dictum rather than as the basis of holding, and even as dictum typically in cases “of gross negligence in which the departure from medical standards is so blatant that it is possible to attribute to hospital administrators’ constructive knowledge of the error in progress.”\textsuperscript{51}

There is also some scattered language suggesting hospitals may have a duty to non-negligently evaluate the quality of the products they provide and/or sell, even though they are typically not liable for defects in those products.\textsuperscript{52} It is possible a future court will actually decide that such a duty exists, in which case it may be possible to apply it to medical AI.

Even if a creative court is open to applying one or more of these direct liability theories to hospitals for torts arising from medical AI,
from a legal system designer point of view these theories leave much to be desired as a way to fit the form to the fuss. Supervision of individual instances of care matches poorly with the opaque and systemic issues that arise from AI errors and place-dependent performance variation. Liability for failure to provide adequate equipment could potentially encompass the reasonable vetting of AI systems prior to implementation, and maybe the ongoing maintenance, monitoring, and tweaking necessary to ensure robust care performance.

The problem with direct liability for hospitals, as the situation stands now, is that even if it applies to hospitals choosing, evaluating, and maintaining AI systems, they lack the information necessary to effectively reduce potential injuries in undertaking those duties. Where the AI system is developed outside the hospital system, these theories particularly miss the mark. The hospital may know general facts—the type of data on which the AI system was trained, approximately how it was trained and validated, and the like—but typically won’t have granular knowledge about exactly what data were used, how exactly the system was trained, what precise parameters were used in validation, and the like. The hospital might lack capacity to analyze these questions, but more fundamentally, developers typically don’t disclose this sort of information, preferring to keep it as a trade secret to protect their product.\footnote{Price II, Big Data, Patents, and the Future of Medicine, supra note 4, at 1432–33.} In fact, even when hospitals do learn things about how AI systems perform, they may be restrained from sharing those facts with other hospitals by gag clauses in the contracts they sign with developers.\footnote{Evans & Pasquale, Product Liability Suits for FDA-Regulated AI/ML Software, in The Future of Medical Device Regulation: Innovation and Protection, supra note 34, at 28.}

At least as things stand now, applying direct liability to hospitals for failures of their AI systems will not solve the question of how they are supposed to effectively minimize injuries caused by those systems.

These problems are less pernicious where AI systems are developed within or with significant cooperation of a hospital system. It is more reasonable in such cases to expect the hospital to know much more about the data sources, validation process, testing parameters, etc., and to fault it for poor supervision or testing before deployment on patients. One could potentially foresee a fork in the path wherein direct liability theories as to medical AI are developed and applied against hospitals for in-house medical AI development, but not against hospitals that adopt products developed outside the hospital. Such a bifurcated approach would be messy, given that in a particular hospital or even a particular patient’s episode of care both kinds of medical AI might be used. It might also lead to a dynamic where hospitals are given a
liability incentive to use externally developed medical AI, which actually heightens the risk of contextual bias in the AI itself, since it means it was developed on other populations with other arrangements of care teams, scheduling, etc. Apart from the bias issue, if one believes that the best medical AI is likely to be designed to solve real problems faced by a community with the input of members of that community, pushing hospitals through the threat of increased liability to “off the rack” and externally produced medical AI also does not seem great.

The real question, as always, is “as against what?” It may be that the two-track internal/external distinction is superior to a baseline where there is no hospital direct liability for either internal or externally developed medical AI. But it may be inferior to a system that applies enterprise liability to both kinds of medical AI, the approach we turn to now.

B. Enterprise Liability Generally

The direct and vicarious liability theories described above are exceptions to a general rule of hospital non-liability for patient adverse events.

By contrast, enterprise liability, as Kenneth S. Abraham and Paul C. Weiler argued in their canonical 1994 *Harvard Law Review* article on the subject, would make “hospitals liable for all malpractice by their affiliated physicians”—in addition to the other forms of liability they already encounter—which “would better serve the goals of tort law than does the current individual liability regime.”55 The idea of hospital enterprise liability first rose to public policy prominence as part of the Clinton Administration’s proposed health care reform in the Spring of 1993, though the idea travelled under different names as early as the 1970s in some academic circles.56 Abraham and Weiler consider proposals to impose enterprise liability for all patient injuries on hospital systems (“Delivery-based liability”) versus imposing it on insurance companies and health plans (“Financing-based liability”).57 Courts and legislatures have not imposed enterprise liability, despite the push by scholars.58 We focus on the hospital as enterprise approach for injuries associated with medical AI—we remain agnostic about the propriety of financing-based liability for other forms of enterprise liability. Below, we consider an alternative proposal of siting enterprise

---

56. Id. at 384.
57. Id. at 415–20.
58. E.g., Benny Chan, *Applying a Common Enterprise Liability Theory to Clinical AI Systems*, 47 Am. J.L. & Med. 351, 370 (2021). There have been some hospital systems that have voluntarily adopted forms of enterprise liability for themselves. Id. at 370–71.
liability with *developers* rather than hospitals and explain why we think hospitals are better suited as the relevant enterprise. But first the affirmative case for hospitals as the target.

Why focus on hospitals as the locus of enterprise liability when it comes to medical AI? Much of what proponents of enterprise liability have said about medical injury in general and the role of hospitals in forestalling it apply with particular force to medical AI. First, let us rehearse the general case for hospital enterprise liability. As Abraham and Weiler put it nearly thirty years ago:

Viewed along the three dimensions of compensation, administration, and prevention, the hospital is a far better candidate for malpractice liability than are individual physicians. . . . The vast majority of the injuries that result from malpractice stem from treatment delivered under the auspices of hospitals—if only by physicians with the right to admit their patients to the hospital. Hospitals 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 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. In each of these respects, hospitals are in the best position to make judgments about whether any such steps will reduce the risk of injuries to their patients. And some hospitals—particularly the more prominent teaching hospitals—are most likely to conduct serious research and development about ways of enhancing the quality of care in the over-all health care system. To the extent, then, that imposing malpractice liability may influence these cost-benefit judgments about patient safety, hospitals would seem to be the prime candidates for that legal role.

Bill Sage puts it succinctly when it comes to medical errors and adverse events: “The likely solution is reducing fragmentation and improving coordination of care through a greater institutional role in both liability risk-bearing and clinical practice. In other words, the

---

59. Our focus is on hospital enterprise liability for adverse events related to the use of medical AI. Barry Furrow has looked at what is almost the opposite idea—using AI tools to detect and *reduce* adverse events and premising liability on the failure to undertake such system-wide analyses:

The case of detection promised by data analytics points to a new standard of care for hospitals—enterprise responsibility for adverse events—beyond the narrow culpability tests of ordinary tort cases. Enterprise liability focuses on the tools a hospital should deploy to spot patient harms . . . Data analytic tools enlarge a hospital’s ability to track physician adverse events and other measures of substandard care. Data analytic tools spot eruptions of patient adverse events; deep learning versions of data analytics can even anticipate and prevent them. Some courts have talked of negligent supervision and negligent credentialing in terms of an affirmative duty to detect problems among physicians.


solution is enterprise liability." More specifically, enterprise liability arguably shifts liability to the actor that has more ability organizationally and financially to reduce medical errors and gives them the incentive to do so.

On the ability piece, hospitals have a better capacity to see the “big picture” than do individual physicians and to “identify and put into place cost-effective safety improvements, especially when those improvements cross the boundaries of individual physician practice, specialty or department, and the professional divide that separates doctors, nurses, and health-related professions.” This is especially essential given the modern view, captured by the Institute of Medicine’s essential 2001 report *Crossing the Quality Chasm: A New Health System for the 21st Century*, which argues that most medical errors and adverse events are caused by system processes not errors by individual providers. Hospitals—not individual physicians—design, implement, and can ultimately improve such systems. Hospital systems, especially large hospital systems, also have the resources to implement these systems.

On the incentive piece, physicians typically have medical malpractice liability insurance that fully covers them—and that insurance is typically not experience-rated, meaning that incentives to avoid patient injury are dampened. By contrast, hospitals’ insurance costs are more closely tied out to their past payouts such that they have a stronger liability incentive to invest resources to reduce medical errors since such reductions reduce their insurance costs. If properly implemented, enterprise liability thus has more of an ability to align incentives for the least cost avoider (more on that below).

---

62. E.g., Peters, Jr., *supra* note 45, at 376.
63. Id. at 376–77.
65. See Peters, Jr., *supra* note 45, at 379 (“Even when an injury appears to have been caused by an individual error, prevention of similar injuries in the future is often best accomplished by changing the system or environment in which individual doctors and nurses work. For example, systemic changes, like augmenting the size of the workforce during periods of heavy workload and requiring that a second physician confirm an initial diagnosis or test interpretation in pre-selected circumstances associated with high error rates, are promising ways of preventing individual errors and protecting patients from common sources of injury. By contrast, individual physicians possess much less capacity than hospitals to recognize weaknesses in the overall system of delivery and to correct those weaknesses, thereby reducing errors and injuries. Of course, this is the central claim of modern patient safety advocates. As a result, most experts believe that hospitals are better situated to improve patient safety than individual physicians.”).
C. Why the Case for Enterprise Liability Applies with Force to Medical AI

The theoretical arguments for enterprise liability generally apply with force to the case of hospital use of medical AI. We say “theoretical” because we readily concede that it is hard to know for sure how these ideas will play out—theory can only take us so far and at some point there must be a leap and a handoff to empirical study. 68

Just as with personnel, hospitals “select[] and provid[e]” the medical AI used by physicians and nurses. They, in a sense “hir[e] and fir[e]” the medical AI that is used in their decisions about which medical AI to acquire or license. 69 And some hospitals—particularly the more prominent teaching hospitals—are most likely to conduct serious research and development about ways of enhancing the quality of care through medical AI and how to best integrate it. Now one might say much of this is also true about an MRI machine or a set of masks for infection control—a hospital system must choose a vendor, can stop using a vendor when problems emerge, and academic medical centers might evaluate how much the item improves patient care and in what circumstances. Fair enough, but what is important about medical AI is the bespoke way it (when done properly) has to be integrated into a medical workforce—what one of us has called the “system view.” 70

As discussed above, the same AI system may improve performance dramatically with the right care team, with the right training, with the right supervision, and the right procedures in place for human in the loop review in some instances, but will be a disaster when one or more of those things change.

68. Even at the level of theory there are many system design choices apart from whether to endorse enterprise liability that will help determine what actually happens. For example, if the economic cost of bringing a lawsuit is too high the effect of enterprise liability in this context will be blunted. That may push us towards adopting a fee-shifting regime or other solutions. The likelihood of bringing suit in cases where adverse events occur may also be mediated by health-care providers’ transparency about the use of AI and patient feelings about the members of their care team and how disclosure of error was handled. An emerging but sophisticated literature considers hospital use of communication-and-resolution programs, for example, Thomas H. Gallagher, Michelle M. Mello, William M. Sage, Sigall K. Bell, Timothy B. McDonald, & Eric J. Thomas, Can Communication-And-Resolution Programs Achieve Their Potential? Five Key Questions, 37 Health Affairs 1845 (2018), and it would be interesting to examine how they intersect with AI-related enterprise liability regimes to either enhance or blunt tort laws’ effects. These are important questions (among many others), but we focus here more directly on the argument for enterprise liability simpliciter, acknowledging that even if we persuade you, dear reader, there are many more details to work out.

69. See supra text accompanying note 60.

70. Gerke et al., supra note 8; see also Price II, Distributed Governance of Medical AI, supra note 6, at 16.
It is the hospital control and supervision over the acquisition and evaluation of medical AI, the evaluation and control over the staffing of the humans who work at the hospital, and the oversight, design, and evaluation over the interaction of the two that makes this such a good case study for the benefits of enterprise liability. Thus, the optimal implementation of medical AI to improve patient care very much transcends the “boundaries of individual physician practice, specialty or department, and the professional divide that separates doctors, nurses, and health-related profession.”\textsuperscript{71} The fragmentation in control and supervision also produces a corresponding “responsibility gap whereby it is difficult to assign blame to any one party.”\textsuperscript{72}

Hospital coordination is especially crucial given the time horizon involved, in two separate ways. First, some of the changes may be costly upfront but pay dividends going forward—something hospitals can take into account but physicians, for example, cannot. Second, there is substantial value in so-called “adaptive” as opposed to “locked” medical AI that learns as it encounters more patients in the actual setting of deployment, but this creates a tough monitoring problem for those overseeing the AI. Who can determine when an AI system’s operation is different enough that additional review is required, and who can conduct that review? The hospital is in a far better position than anyone else.\textsuperscript{73}

To be sure, convincing as one might find the theoretical arguments for hospital enterprise liability, they have made relatively little headway in practice. Nevertheless, the political economy of enterprise liability may be more feasible to stakeholders in the more bounded space of medical AI—and our goal here remains to imagine the “ought” rather than to predict the “will.”\textsuperscript{74}

\textsuperscript{71} Peters, Jr., \textit{supra} note 45, at 376–77.

\textsuperscript{72} Chan, \textit{supra} note 58, at 374.

\textsuperscript{73} One of has discussed this issue, what we term the “update problem,” in the context of the difficulty of a regulator like FDA. \textit{See generally} Boris Babic, Sara Gerke, Theodoros Evgeniou & I. Glenn Cohen, \textit{Algorithms on Regulatory Lockdown in Medicine}, 366 Scir. 1202 (2019). It is very hard to imagine an individual physician undertaking any of that kind of monitoring. It is more plausible for large hospital systems to do it. Even here though, the standard of care would have to be carefully calibrated before liability is imposed for failure to detect significant changes in an algorithm’s performance. There are some delicate questions of whether, for example, FDA approval (in cases where the medical AI has been reviewed) should act as a partial shield, how to parcel out liability for failure to detect performance changes between hospital and developer, etc. What makes it tricky is that there is a lot of benefit to adaptive over locked algorithm, including in reducing contextual bias, and too much liability risk may cause hospital systems to stick to the locked versions.

\textsuperscript{74} One straightforward reason our proposal may be more promising, from a political economy perspective, than traditional enterprise liability proposals is because it seeks to make a much smaller change in terms of which stakeholders’ oxes might be gored. Second, at the time robust enterprise liability proposals first entered real discussion, there was concern by physicians that
IV. DELINEATING THE BORDERS OF ENTERPRISE LIABILITY

Enterprise liability, of course, could rest on either hospitals or developers, but we propose to rest it, at least initially, on hospitals.75

As we have mentioned, neither hospital nor developer fits neatly into the role of cheapest cost avoider, because each has a subset of the information and capabilities necessary to ensure that AI systems are safe and effective as used in a particular place.

So why put the burden of enterprise liability on hospitals rather than developers? We propose three reasons: close patient connections; encompassing physician malpractice; and administrative simplicity.

First, patient connections. While developers have the most knowledge about how the AI system was created, hospitals ultimately have the most information about their own patients, procedures, workflow, and capabilities. Hospitals also have the greatest capacity to systemically monitor performance in the course of ordinary use, and to maintain and improve AI systems to account for changes in patient populations, illnesses encountered, or hospital capacity. To the extent that some hospitals lack that capacity, they are best suited to know that and acquire it.76 Ultimately, hospitals are where the patients are, and where they receive care.

Second, encompassing physician malpractice. While hospitals are the site of the systemic evaluation most appropriate for seeing how well the AI performs in a particular place, physicians are ultimately going

---

75. Benny Chan, inspired by David Vladeck’s proposal for imposing common enterprise liability on the manufacturers of components of autonomous vehicles, has proposed a still wider conception of enterprise liability where “physicians, manufacturers of clinical AI systems, and hospitals that employ the systems [should be] considered to be engaged in a common enterprise for the purposes of tort liability.” Chan, supra note 58, at 352 (applying David Vladeck, Machines Without Principals: Liability Rules and Artificial Intelligence, 89 WASH. L. REV. 117, 149 (2014)). For the reasons we discuss in this section, we are less sure about expanding the boundaries to encompass all of them in a single doctrinal structure, but we confess that our thoughts are very tentative here and it could be that we are too conservative in the scope we suggest.

76. The lack of capacity is not particularly amenable to intervention by the tort system except by incorporation into what counts as “reasonable” under a negligence standard, should negligence remain the applicable standard. But hospitals could hire outside experts (perhaps a good thing!) to help—in some cases, even experts provided by medical AI companies themselves, in the mode of medical device manufacturers who send trainers or monitors along with their complex and expensive new devices.
to be involved in instances of individual care and injury. To the extent that it is confusing to determine exactly what injuries arise from individual misuse of AI systems as something systemic, locating liability in the hospital addresses this confusion, at least from the perspective of centralizing incentives and centralizing responsibility from an injured patient’s point of view. This allocation might have been more wrenching in an earlier era when most physicians involved in hospital care were independent contractors. In a time when many physicians are employed, and hospitals are liable for the actions of many others through various doctrines of vicarious liability, localizing AI-related liability purely in hospitals represents less of a sea change. And where some theories of direct hospital liability have turned on the “care coordination” role of hospitals, the implementation of a centralized AI system to guide care perhaps epitomizes how hospitals fill that care-directing role.

Third, administrative simplicity. It may be quite difficult to disentangle various systemic-level questions about when AI failures result from inadequate training, nonrepresentative data, improper implementation, insufficient integration into clinical workflows, failure to monitor adequately, or any of the other tasks that go into maintaining a well-functioning AI system in a particular environment. Localizing liability also helps put tort incentives all in one place, rather than scattering them with the risk of uncertainty-driven ineffectiveness. And of course there’s little reason to believe that patients have any ability to know who is involved in opaque behind-the-scenes AI failures, as between hospitals and developers. To the extent that they expect anyone to take charge, it’s most likely the hospital providing their care.

For these reasons we think that the default allocation of enterprise liability for patient injuries connected to medical AI should be to the hospital system and not the developer. But as we hinted above, we might want this default to shift under certain circumstances—we focus on access to information, though others may delineate other cases for such shifts.

A. A Default Shift Relating Access to Information

While our focus is on tort law, it is useful to take a sidestep for a moment into contract law theory. In gap-filling in a contract, one approach is to set a default against the interests of the more informed...
party in the interaction, to force them to reveal private information in
the process of contracting around the default. A related question is
how to set the “altering rule,” the rule that tells “private parties the
necessary and sufficient conditions for contracting around a default.”
The same idea has a cousin in civil procedure and in employment law
with the concept of burden shifting as well as in tort law itself in doc-
trines like res ipsa loquitur (“the thing speaks for itself”), wherein some
injuries are so obviously the result of negligence that negligence can be
assumed even if the plaintiff lacks information to show it specifically.

Our recommendation of hospital enterprise liability is best thought
of as a default rule, but one whose altering rule (rather than the setting
of the default itself) is designed to incentivize the sharing of informa-
tion, in particular information about the medical AI system itself. Medi-
cal AI developers, like the developers of AI more generally, have been
highly secretive regarding their products. In part this is because of the
underlying intellectual property treatment of AI at least in the United
States: often the algorithms used in medical AI themselves are not pat-
etable subject matter or fail to meet one of the other requirements of
the Patent Act. As a result, instead of a regime that requires disclosure
as the price of exclusivity, we have ended up in a regime that relies much
more heavily on trade secrecy, including of algorithms, validation tech-
niques, and underlying data to protect investment in AI development.

But our case for hospital enterprise liability—relying on the hospi-
tal’s superior position to be able to test, shop for, monitor, and imple-
ment medical AI—is in tension with settings where the developer’s
unwillingness to share information makes it difficult to do so effectively.
It would be strange to penalize hospitals for failing to use their superior
information to improve safe implementation when they lack access to
exactly that information.

For this reason, we think it is worth exploring an altering rule that
shifts enterprise liability from hospitals to developers in cases where
the developer holds back things like training data sets and parameters

78. Ian Ayres & Robert Gertner, Filling Gaps in Incomplete Contracts: An Economic Theory of
79. Ian Ayres, Menus Matter, 73 U. CHI. L. REV. 3, 6 (2006). There is a similar idea of a “trans-
formation rule,” a rule that specifies the “conditions for transformation of the legal arrangement with
respect to either . . . the brand of legal entitlement or its enforcement rule.” Oren Bracha, Standing
Copyright Law on Its Head? The Googlization of Everything and the Many Faces of Property, 85
80. Price II, Big Data, Patents, and the Future of Medicine, supra note 4, at 1420; but see Mateo
Aboy, Seth Raker & W. Nicholson Price II, Mapping the Patent Landscape for Medical Machine
Learning, 41 NATURE BIOTECH. 461, 466 (2023) (noting the increasing patenting of medical AI
systems).
81. Price II, Big Data, Patents, and the Future of Medicine, supra note 4, at 1407.
LOCATING LIABILITY FOR MEDICAL AI

use, any explainability algorithms they run, etc., that actually would enable the hospital to reasonably evaluate and fine tune the model. In such instances, as an information-forcing device, we recommend that the default be reversed to impose enterprise liability for these injuries on the developer. To be sure the devil will be in the details—exactly how much information disclosure or lack thereof is enough to trigger the altering rule would have to be worked out, either statutorily or more likely by common law judging.

Now one might argue that this is a problem, but it is a problem for the parties not for the courts or tort law. That is, hospital systems could always choose only to acquire medical AI from developers who share the requisite information, and there is no need for a legal doctrine here; hospital systems are sophisticated parties and if they failed to require additional information from the developer, why not take that as the result of private bargaining over price, implementation, licensing, etc.? Who is the court to order a different allocation than the one the parties reached?

It is possible that this pushback is right, but we have a few reasons why we are still a bit skeptical. First, while the parties are sophisticated, this is an area where it may be hard to know what one does not know—that is, hard for hospital systems to correctly identify (and potentially insure against) the scope of their enterprise liability when it is the interaction of an AI system of which they may not have full information with the integration into and usage by a hospital workforce and system that they cannot fully anticipate before it is actually implement. Otherwise put, if you were to make one of us (on the sophisticated side of medical AI knowledge) in charge of negotiating with the developer and making trade-offs between liability, information sharing, price, etc., we would find the task daunting.

Second, the system may have interest in pushing towards information disclosure even when the parties might not bargain for it. We tend to think hospital systems are much more likely to be able to harness medical AI and learn from errors than developers, in part, because so much of the value depends on the implementation. Doing so depends on having more, not less, information about the medical AI. If that is true, it may be worthwhile for the liability doctrines to reinforce that flow of information rather than remain agnostic about it. It may also be socially useful—outside the context of injury-reducing implementation—to promote the sharing of information that can help facilitate innovation, competition, and regulatory oversight.82

Finally, from a litigation perspective there is a real risk of whack-a-mole, or if you prefer, whipsawing. A patient, knowing only that they are injured, will have quite a lot of difficulty parsing out whether the injury is a result of the developer’s design choices, the hospital’s implementation, or more likely the interaction of the two. It may be easier from a litigation perspective (thinking among other things about discovery) to sue a hospital system that has the information disclosed to it, then suing the hospital system without such information disclosed or suing both the hospital and developer who can point fingers at one another. Plainiffs and the court system may benefit from a single duty bearing entity to sue rather than parceling things out between the two defendants.83

B. Some Lingering Issues

As befits the genre of the “symposium piece,” we have raised more questions than we have answered and floated some ideas with some arguments behind them but not purported to fully defend them. Indeed, we are not ourselves completely convinced of the right answers here and view our role as opening the conversation rather than closing it. With that spirit in mind, here are a set of additional open questions that we see.

The “Geographical” Scope of Enterprise Liability for Medical AI: We have been operating, at least implicitly, with a mental model of the modal liability situation as involving a medical AI deployed at a hospital that causes injury to the patient at the hospital. But as with debates about enterprise liability more generally, beyond that case things get murkier. Scholars have raised questions about whether enterprise liability for the entire malpractice issue. They argue:

Adoption of [Enterprise Medical Liability] (EML) would almost certainly reduce the costs of administering medical malpractice litigation, because EML would eliminate multiple individual defendants. When each defendant has separate counsel, legal costs are bound to be substantially higher, as are total malpractice insurers’ expenditures for monitoring the case and otherwise participating in it. In addition, whenever arriving at a settlement depends on securing the agreement of each of a number of defendants, the prospect of settlement is reduced, and additional litigation costs are incurred. EML would likely reduce both sets of administrative costs. Because individual physicians would not be defendants under EML, the costs of multiple representation would be eliminated. Similarly, because the responsible enterprise would be the sole defendant, the additional litigation costs that are sometimes incurred when some, but not all, defendants are willing to agree to a global settlement would also be eliminated. Although we cannot systematically quantify these cost savings, we estimate that they could easily amount to twenty to thirty percent of the current cost of malpractice insurance.

Abraham & Weiler, supra note 55, at 406.
liability should extend, for example, to “injuries occurring in outpatient facilities or those caused by errors during office visits following hospitalization should be included.”\textsuperscript{84} The AI equivalent of the question might extend not just to these settings but also questions about the use of adjunct technologies that collect data or act on hospital instructions—cough monitors, insulin pumps, etc.\textsuperscript{85}

The **No-Fault Alternative**: Enterprise liability requires proving that a hospital system has failed to meet its standard of care and that this caused the injury in question. A different possibility, inspired by worker’s compensation or vaccine compensation funds, would be to adopt a no-fault system entirely.\textsuperscript{86} Patients who can demonstrate injury from the use of medical AI could recover a set amount without showing fault, potentially administered by a government body or another third party, with the cost of such a fund paid for by some combination of fees/taxes imposed on hospital system (in proportion to their usage of medical AI) or on the developers. Might this be even better than enterprise liability? Whether better or worse, how does the political economy around such a proposal compare?\textsuperscript{87}

The Relation of Hospital AI Enterprise Liability to Physician Liability: Should the imposition of enterprise liability on hospitals for the use of medical AI always preempt liability for physicians? It would seem that there are cases where a hospital system adequately met its duties, and yet the physician’s conduct fell below the standard of care in using the medical AI. For example, to use a very extreme hypothetical, imagine the AI system picks out a particular radiological scan as one that is inconclusive and requires a radiologist to read, and the radiologist flips a coin in response. There are also hypotheticals we can construct along the lines of what we have argued above where physician liability seems unfair and inappropriate: for example, a physician relies in part on the advice of a medical AI in determining whether to recommend watchful waiting or a prostatectomy where the physician has every reason to believe improves accuracy but the hospital system, had it non-negligently investigated from its superior vantage point overseeing use across many patients, would know produces bad results for a small subset of patients with particular characteristics. Other cases, though, are more on the precipice. Consider whether one should be able to

\begin{itemize}
\item \textsuperscript{84} E.g., Peters, Jr., supra note 45, at 390.
\item \textsuperscript{85} See, e.g., David A. Simon, Carmel Shachar & I. Glenn Cohen, Unsettled Liability Issues for “Prediagnostic” Wearables and Health-Related Products, 328 JAMA 1391 (2022).
\item \textsuperscript{87} Others have considered keeping it in the tort system but imposing strict liability rather than a fault-based alternative. See Chan, supra note 58, at 378–82.
\end{itemize}
sue both the hospital and the physician in instances where the physician did not “override” an AI recommendation, but the hospital system designed the AI system’s integration into care delivery in such a way that authorizing such an override was cumbersome and required multiple reviews by other physicians. A somewhat separate but interesting case involves “off label” use by physicians of AI systems that have been appropriately cleared by the hospital system for other uses. One might be tempted to think that this (by which we mean liability) is on the physician, but one’s intuition might shift in a case of foreseeable misuse. To the extent that hospitals are vicariously liable for physician malpractice, of course, these issues will tend to collapse (at least from the hospital’s point of view, if not the physician’s).

Insurance and the Cross-Subsidization of Riskier Specialties: In some of the most important writings about adopting hospital enterprise liability for medical malpractice in general one argument made in its favor is that it would help cross-subsidize and average out the insurance costs between physicians in specialties with higher liabilities (like obstetrics and neurosurgery) and lower ones. It is worth thinking about how this dynamic plays out with medical AI for different kinds of medical specialties. Is the effect of adopting medical AI enterprise liability for obstetrics more “powerful” or “disruptive” than the same for orthopedics? Within a hospital system, how should the accounting/cost-sharing take place? Should the cost of enterprise liability for medical AI be “spread” across departments or should those that benefit more from the adoption “pay in” more? Finally, and this relates to our discussion below as to mixed cases where the injury results both from medical AI and non-AI, how does shifting only the AI-related portion to enterprise liability affect incentives for AI adoption? If, for example, liability costs are particularly high in obstetrics, does that give that specialty more of an incentive to adopt medical AI because more of the liability of individual practitioners is now shifted to the cross-subsidized enterprise space? If so, is that a bug or a feature?88

The Contractual Overlay: Very early on (it was several thousand words ago so we do not fault you if you have forgotten!), we noted that Digital Diagnostics is one of the few companies for which it has been publicly reported that it contractually indemnifies providers that use its autonomous IDx-DR diabetic retinopathy detection system (and carries its own malpractice insurance).89 To what extent is such private contractual rebalancing of liability risk between developers and hospital

---

88. E.g., Peters, Jr., supra note 45, at 387–88.
89. Abrámoff et al., supra note 7, at 139.
systems a good thing?\textsuperscript{90} Imagine our proposals for enterprise hospital liability were adopted. To what extent should we treat that as a default versus immutable allocation of liability—that is, should the law have rules about when the liability can be reallocated towards a developer?\textsuperscript{91} Again, one might initially react that these are sophisticated private parties; why should the law step in? But of course, there are externalities for patients. If one believes the arguments for enterprise liability we have offered above, siting liability with the hospital system is more likely to prompt a reduction in errors and injuries.

\textit{Mixed AI and non-AI Caused Injuries:} Our proposal has been modest in that it recommends hospital enterprise liability for injuries stemming from the use of medical AI and not the \textit{general} adoption of hospital enterprise liability. In so doing, though, it puts a lot of pressure on the categorization. Medical injuries often have several overlapping causes and part of the design of safe systems is to build in redundancies that catch and correct an error, such that when an injury occurs there may have been multiple points of failure. What happens when some of those points of failure in a single care episode are AI-related and others more AI-unrelated? It seems a bit odd to have litigation proceed along two tracks for the same injury where the AI-related one is treated as an instance of enterprise liability while the other has only available or vicarious liability theories. It is true that patients sometimes offer two or more separate theories of liability (such as strict liability and negligence) for the same injury, but this seems somewhat different. One solution would be to develop a center-of-gravity kind of test to determine which “system” of liability the entire case belongs in, but specifying exactly which factors matter is challenging. (How much of a cause of the injury the AI was? How egregious was the deviation from the

\textsuperscript{90} Hospitals may also indemnify physicians, or vice versa.

\textsuperscript{91} While we have acknowledged the possibility of contractually re-allocating liability between the AI developer and the hospital system, what about re-allocating between the patient and the hospital system? That is, should the availability of enterprise liability as a cause of action for the plaintiff against the hospital be treated a default rulewaivable by patients? U.S. courts have typically rejected the advance waiver of medical malpractice liability. See, e.g., Tunkl v. Regents of Univ. of Cal., 383 P.2d 441, 442 (Cal. 1963) (en banc); Nadia N. Sawicki, \textit{Choosing Medical Malpractice}, \textit{93 Wash. L. Rev.} 891, 893 (2018) (noting the general position); \textit{id.} at 910–13 (noting critiques and rebuttals). Without staking out an explicit position on this question, we think the case for allowing waiver in the case of medical AI errors is considerably weaker than in the typical malpractice case because the information deficit faced by patients seems much greater: As hard as it is for a patient to evaluate a waiver of typical medical malpractice liability, it seems much harder for a patient to effectively evaluate the equivalent waiver for a hospital system’s liability regarding use of a medical AI system, where the patient lacks good information on how the AI system is used in care, what review the AI system has undergone before implementation, and how the AI system works.
standard of care for the AI and non-AI cause? Etc.) Moreover, any attempt to sort at the front end might add cost (at worst a mini-trial).\(^{92}\)

None of these issues are insurmountable, but they are all challenging. Nevertheless, if our argument succeeds and you, dear reader, come to believe that enterprise hospital liability for injuries from medical AI is desirable, there are questions that must be faced.

**Conclusion**

Tort law is murky but useful. Where medical AI is involved, tort law’s murkiness blends with AI’s opacity and differential performance to create real problems in setting incentives for parties with the ability to avoid harm. Making liability clearer, and setting it initially on the health systems most directly responsible for coordinating, has the potential to improve both adoption and safety of AI systems in medicine.

---

92. Looking further along, we might imagine that rules applied to the context of medical AI could spread into other areas of healthcare liability. Would the difficulty of determining that an injury resulted from AI issues, especially as AI spreads across medical care, mean that more and more injuries would be subject to hospital enterprise liability, at least at a first cut? We do not mean our proposal to be a Trojan horse, but we also do not suppose this outcome would be so terribly bad.