

---

---

# LIABILITY’S BLIND SPOT: HOW HEALTHCARE AI CAN EARN PATIENTS’ TRUST

Shih-wei Chao\*

*Imposing liability is often society’s first response to the risks and harm accompanying new technologies. As Artificial Intelligence (“AI”) gains a foothold in the healthcare sector, liability is thus the likely focal point when addressing its potential dangers, such as misdiagnoses or robots injuring patients. One supposed benefit of liability regimes is that they help earn patients’ trust, a paramount objective in healthcare, by incentivizing safety. This Article contends, however, that liability cannot build trust. Liability, at best, promotes trustworthiness, which is different from trust. At worst, liability even erodes trust. This blind spot of liability regimes suggests that earning trust may require a broader, industry-wide effort.*

## I. INTRODUCTION

In Taiwan’s Taichung Veterans General Hospital, a white humanoid robot strolls through a hallway. Its face screen displays a smile and a pair of round, dreamy eyes that blink periodically. It stops at a corner to let a nurse pass before proceeding to its destination, a patient’s bedside. A lid on its base compartment opens, and a container holding several vials pops up. The robot lifts the container with its arms, then gently places it on a nearby table. Meet Nurabot, an Artificial Intelligence (“AI”)-powered robot that delivers medication throughout the hospital, promising to enhance efficiency and offset staff shortages.<sup>1</sup>

Nurabot is just one example of AI’s growing role in the healthcare sector. As healthcare demand trends up, AI applications like medical image analysis, service robots, and even surgical assistance are becoming increasingly necessary. But AI comes with a caveat. Due to AI’s “black-box” nature, patients constantly have to worry that AI will commit errors or cause unexpected harm, making trust

---

\* Assistant Professor, Department of Financial and Economic Law, National Chung Cheng University, Taiwan. J.S.D., University of California, Berkeley. I am grateful to the participants of the Seventh Taiwan-Vietnam Law Forum: International Conference on Liability in the Health Sector for helpful comments and suggestions. All errors are my own.

1. NVIDIA, *Foxconn Builds Robotics for Healthcare with NVIDIA Physical AI* (YouTube, May 19, 2025), <https://www.youtube.com/watch?v=3YbyqaV0CDI> [<https://perma.cc/5D7S-VZYG>].

difficult to achieve.<sup>2</sup> This lack of trust is unfortunate because in the healthcare sector, trust carries both intrinsic value and extrinsic benefits.<sup>3</sup>

To foster patients' trust in AI, an intuitive starting point is to impose liability regimes holding medical professionals, hospitals, or AI companies responsible. As the reasoning goes, the threat of liability makes AI safer, and safer AI earns trust.<sup>4</sup>

This Article contends, however, that such reasoning is misguided. At best, liability only creates *trustworthiness*, but not trust. At worst, liability can even erode trust. The common liability-centric mindset fails to appreciate the nature of trust, and the difference between trust and trustworthiness.<sup>5</sup>

Part II of this Article sets forth the meaning of trust and why the healthcare sector should strive to earn patients' trust. Part III discusses why patients might hesitate to trust AI when it participates in care. Part IV outlines liability regimes and the rationale that they foster trust. Part V argues that liability is insufficient to create trust by distinguishing between trust and trustworthiness. Part VI suggests alternative approaches to genuinely build trust.

## II. TRUST IN HEALTHCARE AND WHY IT MATTERS

In the healthcare sector, most would agree that it is important for patients to trust the medical professionals and institutions providing care. But what is trust, after all? And why exactly is trust valuable? In the AI era, when healthcare is drastically transforming, these fundamental questions are more relevant than ever.

### A. Trust

Obviously, it is impossible to perfectly define trust in these few pages, as doing so could easily occupy a lifetime of study.<sup>6</sup> For this Article's purposes, however, a general conception will suffice. Trust entails a sense of vulnerability, where the one who trusts (the "trustor") relies on the one who is trusted (the

---

2. E.g., Carlo Giovine & Roger Roberts, *Building AI Trust: The Key Role of Explainability*, MCKINSEY & CO. (Nov. 26, 2024), <https://www.mckinsey.com/capabilities/quantumblack/our-insights/building-ai-trust-the-key-role-of-explainability> [<https://perma.cc/MC26-SXZ4>]; Mark Bailey, *Why Humans Can't Trust AI: You Don't Know How It Works, What It's Going to Do or Whether It'll Serve Your Interests*, CONVERSATION (Sept. 13, 2023 8:29 AM EDT), <https://theconversation.com/why-humans-cant-trust-ai-you-dont-know-how-it-works-what-its-going-to-do-or-whether-itll-serve-your-interests-213115> [<https://perma.cc/G7HZ-V9BR>].

3. See Mark A. Hall, *Law, Medicine, and Trust*, 55 STAN. L. REV. 463, 477–82 (2002) (identifying trust's intrinsic value and extrinsic benefits).

4. Cf. Zach Harned, Matthew P. Lungren & Pranav Rajpurkar, Comment, *Machine Vision, Medical AI, and Malpractice*, HARV. J.L. & TECH. DIG. 1, 2–3 (2019), <https://jolt.law.harvard.edu/digest/machine-vision-medical-ai-and-malpractice> [<https://perma.cc/5UR5-47WC>] (indicating that liability is one of the first concerns when it comes to introducing AI into healthcare).

5. Hall, *supra* note 3, at 487 (distinguishing between trust and trustworthiness).

6. See, e.g., Carolyn McLeod, *Trust*, STAN. ENCYCLOPEDIA PHIL. (last updated Aug. 10, 2020), <https://plato.stanford.edu/entries/trust/> [<https://perma.cc/Q9LB-MYAG>] (collecting prominent philosophical works on the topic of trust).

“trustee”) to act or deliver on something.<sup>7</sup> In healthcare, patients are inherently vulnerable, relying on physicians and hospitals for treatment. Despite this reliance, paradoxically, trust requires the possibility of things going wrong. When the performance is guaranteed, it makes little sense to speak of trust.<sup>8</sup>

In simple terms, trust means the trustor believes the trustee will follow through, even though the trustee might in fact not live up to the trust. Crucially, trust is a subjective phenomenon taking place purely in the mind of the trustor.<sup>9</sup> In healthcare, each patient may place a different level of trust in the medical institution or the physician.

Because trust is subjective, it can be difficult to ascertain what factors affect trust, and how to effectively build trust.<sup>10</sup> A patient might trust a physician for reasons unrelated to expertise, such as attentiveness, confidence, or a welcoming care environment. Proven safety and scientific reliability might also inspire trust. But none of these influences guarantee trust; its subjectivity means patients may trust or not trust for widely varying reasons.

### B. *Why Trust Matters*

Even though trust can appear nebulous, it remains an essential goal in the healthcare sector.<sup>11</sup> One reason is its intrinsic value, being a virtue in and of itself. The patient’s relationship with the medical professional is often intimate and personal due to the patient’s vulnerable state, and healthcare by nature involves the treatment of the deeply private human body.<sup>12</sup> The level of the patient’s trust embodies how physicians, nurses, and hospitals manage such a delicate relationship. Healthcare where the patient trusts is healthcare done right.

A second reason to pursue trust lies instead in its extrinsic or instrumental benefits of improving treatment outcomes.<sup>13</sup> The most prominent example is probably the placebo effect: when patients trust their physician to provide

---

7. Annette Baier, *Trust and Antitrust*, 96 ETHICS 231, 235 (1986).

8. Christian Budnik, *Can We Trust Artificial Intelligence?*, 38 PHIL. & TECH. 1, 4–5 (2025), <https://link.springer.com/article/10.1007/s13347-024-00820-1> [<https://perma.cc/9JPY-UD98>].

9. Mark A. Hall, *Can You Trust a Doctor You Can’t Sue?*, 54 DEPAUL L. REV. 303, 305 (2005); Daniel Hult, *Creating Trust by Means of Legislation—A Conceptual Analysis and Critical Discussion*, 6 THEORY & PRAC. LEGIS. 1, 10 (2018).

10. See generally Kevin Anthony Hoff & Masooda Bashir, *Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust*, 57 HUM. FACTORS 407, 410, 413 (2015) (recognizing that trust involves irrational emotions, and attempting to distill its components); Kent Grayson, *Cultivating Trust Is Critical—and Surprisingly Complex*, KELLOGGINSIGHT (Mar. 7, 2016), <https://insight.kellogg.northwestern.edu/article/cultivating-trust-is-critical-and-surprisingly-complex> [<https://perma.cc/L5H6-AHPJ>] (highlighting the subjectivity and irrationality of trust).

11. See Hall, *supra* note 9, at 1.

12. See Hall, *supra* note 3, at 477–78 (describing how patients’ vulnerability gives rise to considerable intimacy); Carleen M. Zubrzycki, *Privacy From Doctors*, 39 YALE L. & POL’Y REV. 526, 549 (2021) (commenting on the intimate and personal nature of doctor-patient interactions).

13. E.g., Carole A. Robinson, *Trust, Health Care Relationships, and Chronic Illness: A Theoretical Coalescence*, 3 GLOB. QUALITATIVE NURSING RSCH. 1, 1–2 (2016), <https://journals.sagepub.com/doi/10.1177/233393616664823> [<https://perma.cc/V7AG-TXMJ>]; Johanna Birkhäuer et al., *Trust in the Health Care Professional and Health Outcome: A Meta-Analysis*, 12 PLOS ONE 1, 1 (2017), <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0170988> [<https://perma.cc/9TWH-GWFS>].

effective medication, that alone can produce observable therapeutic benefits, even if the medication was actually a “placebo.”<sup>14</sup> Moreover, a trusting patient might send positive signals that help medical professionals respond with greater confidence and ease, ultimately improving treatment quality.

### C. *Trust in AI Healthcare*

Even when care involves AI or is entirely provided by AI, trust remains vital.<sup>15</sup> Intrinsicly, earning the patient’s trust should still be a core objective of healthcare, since the patient is just as vulnerable and exposed as when treated by humans. Extrinsicly, trust in AI may continue to produce therapeutic benefits similar to the placebo effect. While it is true that patients’ signals of trust do not affect AI like they might human caregivers, a lack of trust in AI can nonetheless be problematic due to trust’s transitive effects; namely, this lack of trust can extend to medical professionals and institutions more broadly, undermining the entire healthcare sector.<sup>16</sup>

## III. AI IN HEALTHCARE

Unfortunately, a “trust gap” is forming in the AI healthcare space.<sup>17</sup> AI promises to enhance efficiency and quality, all the while reducing workload. But AI is notoriously opaque, making it unpredictable and thus unreliable.<sup>18</sup>

### A. *The Promise of AI*

The healthcare sector is nearing a breaking point. In Taiwan, for instance, specialties with long hours, stagnant wages, and high litigation risk, such as urgent care, pediatrics, and gynecology, see physicians and nurses leaving in

---

14. Hall, *supra* note 3, at 479–80; see PEKKA LOUHIALA, PLACEBO EFFECTS 59–60 (2020) (discussing historical records indicating that patients’ trust facilitate the placebo effect); Martin Bystad, Camilla Bystad & Rolf Wynn, *How Can Placebo Effects Best Be Applied in Clinical Practice? A Narrative Review*, 2015 PSYCH. RSCH. & BEHAV. MGMT. 41, 43 (reviewing literature on the importance of trust in inducing the placebo mechanism).

15. See Saleh Afroogh, Ali Akbari, Emmie Malone, Mohammadali Kargar & Hananeh Alambeigi, *Trust in AI: Progress, Challenges, and Future Directions*, 11 HUMANITIES & SOC. SCI. COMM’NS 1, 1–2 (2024), <https://www.nature.com/articles/s41599-024-04044-8> [<https://perma.cc/GLH5-7LHB>] (suggesting that trust is the key to AI technology diffusion in many areas, including healthcare); Madeline Sagona, Tinglong Dai, Mario Macis & Michael Darden, *Trust in AI-assisted Health Systems and AI’s Trust in Humans*, 2 NATURE PARTNER J. HEALTH SYS. 1, 1 (2025), <https://www.nature.com/articles/s44401-025-00016-5> [<https://perma.cc/5MSL-ACG6>] (insisting trust must underpin the use of AI in healthcare).

16. Thomas P. Quinn, Manisha Senadeera, Stephan Jacobs, Simon Coghlan & Vuong Le, *Trust and Medical AI: The Challenges We Face and the Expertise Needed to Overcome Them*, 28 J. AM. MED. INFORMATICS ASS’N 890, 890 (2021); see Hall, *supra* note 3, at 475 (describing trust’s “halo” effect in healthcare).

17. Shez Partovi, *There’s a Trust Gap in Health-Care AI. Here’s How to Bridge It*, FORTUNE (May 15, 2025, 10:12 AM EDT), <https://fortune.com/article/health-care-ai-adoption-trust-gap/> [<https://perma.cc/9247-948R>]; John Ward, *Bridging the AI Trust Gap*, EY (Aug. 5, 2021), [https://www.ey.com/en\\_ie/insights/ai/bridging-the-ai-trust-gap](https://www.ey.com/en_ie/insights/ai/bridging-the-ai-trust-gap) [<https://perma.cc/Q8BE-UPQJ>].

18. See Afroogh et al., *supra* note 15, at 17.

large numbers, while those staying behind are even more overworked.<sup>19</sup> Meanwhile, an aging population is driving up demand for elder care, further straining the system.<sup>20</sup>

Many have touted AI as a solution.<sup>21</sup> AI apparently analyzes some medical images more accurately and much faster than human physicians, suggesting it could play a meaningful role in diagnosis.<sup>22</sup> It could also assist with patient monitoring by collecting data and identifying abnormalities.<sup>23</sup> Robots like Nurabot can further reduce workload by handling short-range delivery.<sup>24</sup> In the (surprisingly not so distant) future, AI may even participate in surgery.<sup>25</sup> All of this helps reduce workload and expand hospital capacity.

### B. *The Perils of AI*

Despite AI's potential, employing AI in healthcare can be risky. Other sectors have already seen cases of AI malfunction and even harm.<sup>26</sup> Generative AI “hallucinates” inaccurate information.<sup>27</sup> A factory robot in China amusingly erupted into chaos, viciously waving its hands as if attacking nearby workers.<sup>28</sup> Autonomous vehicle crashes have made headlines since AI first took the wheel.<sup>29</sup>

---

19. Alan Y Hsu & Chun-Ju Lin, *The Taiwan Health-Care System: Approaching a Crisis Point?*, 404 LANCET 745, 745–46 (2024); see generally Michael Turton, *Notes from Central Taiwan: Taiwan's Health System Is Going to Erode, Not Collapse*, TAIPEI TIMES (May 5, 2025), <https://www.taipetimes.com/News/feat/archives/2025/05/05/2003836330> [https://perma.cc/KW4V-X9ET] (discussing a Lancet paper retraction controversy where Taiwanese officials rebuked specific statistics but failed to address the paper's core claims that the healthcare system is overburdened).

20. Hsu & Lin, *supra* note 19, at 746.

21. Bing-Hung Shih & Chien-Chun Yeh, *Advancements in Artificial Intelligence in Emergency Medicine in Taiwan: A Narrative Review*, 14 J. ACUTE MED. 9, 9–10 (2024).

22. *Id.* at 10.

23. *Id.*

24. Hon Hai Technology Group (Foxconn) Unpacks Artificial Intelligence Progress at NVIDIA GTC, FOXCONN (Mar. 19, 2025), <https://www.foxconn.com/en-us/press-center/events/csr-events/1557> [https://perma.cc/KU59-D55Y]; Prerna Dogra, *Foxconn Taps NVIDIA to Accelerate Physical and Digital Robotics for Global Healthcare Industry*, NVIDIA (May 18, 2025), <https://blogs.nvidia.com/blog/foxconn-smart-hospital-robot/> [https://perma.cc/3RPQ-DYAN].

25. J. Everett Knudsen, Umar Ghaffar, Runzhuo Ma & Andrew J. Hung, *Clinical Applications of Artificial Intelligence in Robotic Surgery*, 18 J. ROBOTIC SURGERY, 1, 1 (2024); *Reimagining Surgery: The Institute Unveils SuPER, a Leading Surgical Robotics and AI Research Centre*, MCGILL U. HEALTH CTR. (Apr. 30, 2025), <https://muhc.ca/news-and-patient-stories/news/reimagining-surgery-institute-unveils-super-leading-surgical-robotics> [https://perma.cc/8DFC-FYDC].

26. See Thor Olavsrud, *11 Famous AI Disasters*, CIO (Aug. 7, 2025), <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html> [https://perma.cc/8PCG-7JNK].

27. See *id.*

28. Ben Cost, *Violent Humanoid Robot Snaps—Attacks Factory Workers in Wild Video: ‘Went Full Terminator’*, N.Y. POST (May 5, 2025, 11:28 AM ET), <https://nypost.com/2025/05/05/tech/violent-humanoid-robot-snaps-attacks-factory-workers-video/> [https://perma.cc/K583-JKKF].

29. E.g., Bigad Shaban, *While Waymo Not Blamed in Multi-car Wreck, It's the First Fatal Collision Involving a Driverless Car*, NBC: BAY AREA (last updated Jan. 21, 2025, 12:21 PM), <https://www.nbcbayarea.com/investigations/waymo-multi-car-wreck-san-francisco-driverless/3766860/> [perma.cc/S326-TMDP]; Dana Hull & Craig Trudell, *A Fatal Tesla Crash Shows the Limits of Full Self-Driving*, BLOOMBERG (last updated June 12,

AI in healthcare is likewise susceptible to error.<sup>30</sup> AI's medical image analysis is far from perfect, which could translate to misdiagnoses if physicians overly rely on it.<sup>31</sup> Delivery robots like Nurabot could bump into staff or patients. Robots could also inadvertently injure patients during close-contact tasks such as physical therapy or helping them out of bed. Not to mention when AI enters the operating room, even the slightest mistake could be fatal.

The problem with AI errors is that they are often unexplainable, therefore both unpredictable and unpreventable. AI is famously a “black box,” as its immensely complex neural networks make tracing its decision-making process near impossible.<sup>32</sup> Even AI companies admit they do not fully understand the systems they create.<sup>33</sup> Unlike traditional “rule-based” machines, AI can behave in unforeseeable ways, and when that happens, there is basically no way to directly “debug.”<sup>34</sup> Training AI with more data can only help it approach—but never achieve—perfection.<sup>35</sup>

### C. *The Trust Gap*

AI's volatility explains why the healthcare sector still lacks trust in it.<sup>36</sup> Specifically, so long as AI participates in healthcare, patients may worry about misdiagnoses, incorrect medication delivery, or other potential harm. To benefit from AI's contributions, while also preserving trust's intrinsic and extrinsic values, the healthcare sector must find ways to foster patients' trust in AI.

---

2025, 4:40 PM CDT), <https://www.bloomberg.com/features/2025-tesla-full-self-driving-crash/> [<https://perma.cc/9352-BRAJ>].

30. KARIM LEKADIR, GIANLUCA QUAGLIO, ANNA TSELIODIS & CATHERINE GALLIN, EUR. PARLIAMENTARY RSCH. SERV., PE 729.512, ARTIFICIAL INTELLIGENCE IN HEALTHCARE: APPLICATIONS, RISKS, AND ETHICAL AND SOCIETAL IMPACTS 15 (2022), [https://www.europarl.europa.eu/thinktank/en/document/EPRS\\_STU\(2022\)729512](https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2022)729512) [<https://perma.cc/U4JK-PEY8>]; Ryan Tracy & Stephanie Armour, *Medical AI Tools Can Make Dangerous Mistakes. Can the Government Help Prevent Them?*, WALL ST. J. (Dec. 2, 2023, 8:30 AM ET), <https://www.wsj.com/tech/ai/medical-ai-tools-can-make-dangerous-mistakes-can-the-government-help-prevent-them-b7cd8b35> [<https://perma.cc/7EW7-682P>].

31. LEKADIR ET AL., *supra* note 30, at 21.

32. See Marzyeh Ghassemi, Luke Oakden-Rayner & Andrew L Beam, *The False Hope of Current Approaches to Explainable Artificial Intelligence in Health Care*, 3 LANCET DIGITAL HEALTH e745, e745 (2021) (suggesting that contemporary explainability techniques cannot produce satisfactory explanations of healthcare AI's individual decisions).

33. E.g., Maxwell Zeff, *Anthropic CEO Wants to Open the Black Box of AI Models by 2027*, TECHCRUNCH (Apr. 24, 2025, 16:28 PDT), <https://techcrunch.com/2025/04/24/anthropic-ceo-wants-to-open-the-black-box-of-ai-models-by-2027/> [<https://perma.cc/6DST-A82T>].

34. Katharine Miller, *How Do We Fix and Update Large Language Models?*, STAN. INST. HUMAN-CENTERED ARTIFICIAL INTELLIGENCE (Feb. 13, 2023), <https://hai.stanford.edu/news/how-do-we-fix-and-update-large-language-models> [<https://perma.cc/QH46-NJM6>].

35. See Cade Metz & Karen Weise, *A.I. Is Getting More Powerful, but Its Hallucinations Are Getting Worse*, N.Y. TIMES (May 6, 2025), <https://www.nytimes.com/2025/05/05/technology/ai-hallucinations-chatgpt-google.html> [<https://perma.cc/JJM3-CLTT>] (reporting that AI errors like hallucination will “never go away” despite developers' best efforts to improve).

36. E.g., Madeline Sagona et al., *supra* note 15, at 1; Paige Nong & Jodyn Platt, *Patients' Trust in Health Systems to Use Artificial Intelligence*, 8 JAMA NETWORK OPEN 1, 4 (2025) (finding evidence that patients lack trust in healthcare AI).

#### IV. LIABILITY REGIMES ATTEMPTING TO BUILD TRUST

To create this sought-after trust, liability seems to be a straightforward starting point. Imposing liability is often society's initial response to new technology's risks and harm.<sup>37</sup> For other AI applications, such as autonomous vehicles and content generation, much of the legal discussion so far has indeed surrounded liability.<sup>38</sup> This gives a good reason to consider liability as a solution to healthcare AI's "trust gap" problem.

At first glance, liability does appear to promote trust. The logic is simple: whoever is liable has the incentive to make AI safer, and patients will presumably trust safer AI. Put simply, liability assures that the AI is as good and safe as it can be. This argument is nothing new, pervading the literature already.<sup>39</sup>

In healthcare, AI liability regimes could come in several forms. One possibility is to assign liability to the physician or nurse who "uses" the AI, similar to using dangerous tools like scalpels or syringes.<sup>40</sup> Alternatively, under the view that AI is more autonomous, medical professionals or institutions could still bear so-called "vicarious" liability based on them having some degree of control over the AI, akin to how employers can be liable for their employees' actions.<sup>41</sup> Another option would be to impose liability on the AI's developers or manufacturers, essentially a variation of product liability.<sup>42</sup> Finally, a more radical idea is to grant AI legal personhood and hold it personally liable.<sup>43</sup> All these liability regimes more or less have the same objective. Under the threat of liability, whoever is liable will do everything possible to make AI safer for patients.

#### V. WHY LIABILITY CANNOT CREATE TRUST

But the key question is: will patients trust AI more because of the assurance of liability? This Article argues the answer is no. Even if liability does incentivize AI safety—which it sometimes does, but not always<sup>44</sup>—what safety produces is

---

37. See Catherine Sharkey, *Products Liability for Artificial Intelligence*, LAWFARE (Sept. 25, 2024, 8:01 AM), <https://www.lawfaremedia.org/article/products-liability-for-artificial-intelligence> [<https://perma.cc/7UAK-D5X5>].

38. See David C. Vladeck, *Machines Without Principals: Liability Rules and Artificial Intelligence*, 89 WASH. L. REV. 117, 128 (2014) (discussing potential liability assignment for AI in general); Sophia H. Duffy & Jamie Patrick Hopkins, *Sit, Stay, Drive: The Future of Autonomous Car Liability*, 16 SMU SCI. & TECH. L. REV. 453, 454–55 (2013) (identifying the potential need for determining liability in case of an autonomous vehicle accident); Peter Henderson, Tatsunori Hashimoto & Mark Lemley, *Where's The Liability in Harmful AI Speech?*, 3 J. FREE SPEECH L. 589, 626 (2023) (discussing liability when generative AI creates harmful content).

39. See *Loomis v. Amazon.com LLC*, 63 Cal. App. 5th 466, 476 (2021) (explaining that liability serves as an incentive to improve product safety); Steven Shavell, *Liability for Harm Versus Regulation of Safety*, 13 J. LEGAL STUD. 357, 357 (1984) (suggesting that tort liability promotes safety by deterring harmful activity).

40. Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315, 1319–20 (2020).

41. Mihailis E. Diamantis, *Vicarious Liability for AI*, 99 IND. L.J. 317, 319 (2023).

42. See generally Jeffrey K. Gurney, Note, *Sue My Car Not Me: Products Liability and Accidents Involving Autonomous Vehicles*, 2013 U. ILL. J.L. TECH. & POL'Y 247, 257–58.

43. Vladeck, *supra* note 38, at 124–25.

44. See Gurney, *supra* note 42, at 272–74.

trustworthiness, not trust. This Part shows why by illustrating the subtle but significant difference between the two. Moreover, liability not only fails to create trust, but might also undermine it. If so, encouraging patients to trust AI with liability regimes would be counterproductive.

A. *Trustworthiness Is Not Trust*

Unlike trust, which is a subjective phenomenon that focuses on the experience of the vulnerable trustor, trustworthiness is an objective attribute of the trustee.<sup>45</sup> It involves assessing whether someone (or something, such as AI) meets expected standards of competence, safety, or reliability.<sup>46</sup> The popular term “trustworthy AI” illustrates this perfectly. IBM defines trustworthy AI as “artificial intelligence systems that are explainable, fair, interpretable, robust, transparent, safe and secure.”<sup>47</sup> These qualities all refer to the AI system itself, and are (intended to be) measurable through some objective metric or statistic. They are not observed in terms of AI users’ subjective experience. That is, measures of trustworthiness are conceptually distinct from the level of trust.

Of course, trustworthiness and trust are related. One way to look at trustworthiness is as the level of trust others should have in someone.<sup>48</sup> Qualities like safety, reliability, transparency, and explainability are good reasons to trust. But the fact that one should trust does not mean one will. Even with trustworthiness providing a reason to trust, a potential trustee might trust or not trust for other unrelated reasons.<sup>49</sup>

B. *AI Liability Regimes Create Trustworthiness at Best, Not Trust*

Given the distinction between trustworthiness and trust, which does liability actually affect? Liability does indeed incentivize those involved to make AI more reliable and safe: AI companies might enforce stricter quality controls, while medical staff might monitor AI outputs more closely or double-check image analyses, for example. In other words, liability promotes more competent AI.<sup>50</sup> Notice that competence involves some measurement in the objective sense, against either some industry norm, duty of care, or product liability defect standard. This means liability regimes target *trustworthiness*, not trust. Again, while liability-induced trustworthiness is a good reason to trust, it neither guarantees nor necessarily increases trust.

---

45. Hall, *supra* note 3, at 487.

46. Hall, *supra* note 9, at 305.

47. Alice Gomstyn, Alexandra Jonker & Amanda McGrath, *What Is Trustworthy AI?*, INT’L BUS. MACHINES (Oct. 24, 2024), <https://www.ibm.com/think/topics/trustworthy-ai> [<https://perma.cc/6WR2-FWRT>].

48. Hall, *supra* note 9, at 305.

49. *Id.*

50. Guido Noto La Diega & Leonardo C.T. Bezerra, *Can There Be Responsible AI Without AI Liability? Incentivizing Generative AI Safety Through Ex-post Tort Liability under the EU AI Liability Directive*, 32 INT’L J.L. & INFO. TECH. 1, 3 (2024), <https://doi.org/10.1093/ijlit/eaac021> [<https://perma.cc/B2TU-WYG4>].

That is assuming, of course, that liability does in fact enhance trustworthiness.<sup>51</sup> When those liable lack the means or knowledge to carry out the improvement, liability may have little or no effect on trustworthiness at all. This can happen, for example, when even AI companies are uncertain how to fix flaws in the AI “black box.” Therefore, to qualify, liability regimes create trustworthiness *at best*.

Compare patients receiving AI-assisted care to a passenger entering a known defective car. If the car malfunctions and crashes, surely someone would be liable, perhaps the car manufacturer, the mechanic in charge, or whoever recommended the car. But the passenger, knowing the car is defective, is not going to trust the car more because of this assurance.<sup>52</sup> For one, in this situation liability clearly did not make the car sufficiently trustworthy, possibly because the car was too complex or key repair parts were simply unavailable. Sure, liability might have marginally improved the car’s trustworthiness. But the passenger still does not trust it for other reasons (such as worrying the entire ride that things might go wrong).

Likewise, patients will not trust AI merely because liability regimes are in place. At best, liability incentivizes trustworthiness, but that is neither certain nor sufficient. Technical barriers could mean that no one can make AI safer for patients, even with the best intentions, which liability cannot change. And even where trustworthiness exists, patients may as well not trust AI for a host of other reasons—some of which are illustrated below.

### C. *AI Liability Regimes Could Even Undermine Trust*

Even worse than failing to build trust, liability regimes can actively undermine trust. The language of rights and the language of trust tend to move in opposite directions: one who insists on asserting rights effectively signals a lack of trust.<sup>53</sup> In addition, the law’s emphasis on liability could reinforce the perception that AI is flawed, leading patients to question AI’s competence.<sup>54</sup> Relatedly, AI failures such as autonomous vehicle accidents are already receiving disproportionate media coverage, in part due to the potential for enormous damages.<sup>55</sup> Publicizing these incidents can fuel public anxiety and weaken patients’ trust in AI.

---

51. See also A. Mitchell Polinsky & Steven Shavell, *The Uneasy Case for Product Liability*, 123 HARV. L. REV. 1437, 1440–41 (2010) (suggesting that in many instances product liability fails to noticeably increase safety, in part because regulations and market forces already provide sufficient incentives).

52. See Hall, *supra* note 3, at 491–92 (contending that liability does little to assure patients about physicians’ competence); Hall, *supra* note 9, at 306 (surmising that a patient’s trust cannot be restored even with a successful liability lawsuit).

53. Hall, *supra* note 3, at 469 (quoting Richard Sherlock, *Reasonable Men and Sick Human Beings*, 80 AM. J. MED. 2, 3 (1986)).

54. See *id.* at 492.

55. See Trisha Thadani, *Lawsuits Test Tesla Claim That Drivers Are Solely Responsible for Crashes*, WASH. POST (Apr. 28, 2024), <https://www.washingtonpost.com/technology/2024/04/28/tesla-trial-autopilot-lawsuit/> [https://perma.cc/Q94V-N5ZR] (analyzing high-profile lawsuits concerning Tesla’s self-driving technology).

Liability regimes can also erode trust by introducing concerns about the uncertain and open-ended nature of assessing damages.<sup>56</sup> Patients may suspect AI systems are designed not to provide optimal care, but to reduce AI companies' or physicians' exposure to legal risk, also known as "defensive medicine."<sup>57</sup> They might also worry that gaps or ambiguity in liability regimes could leave their harms inadequately compensated. Both scenarios create a climate of suspicion rather than trust.

## VI. TRUST-CREATING SOLUTIONS INSTEAD OF LIABILITY

While liability regimes do not create trust, and can even weaken it, this only means the healthcare sector must find other ways to build patients' trust in AI. This Part proposes two possible pathways: certification schemes or embracing a "just culture." Either way, the point is that earning trust should be an industry-wide, multi-stakeholder effort. Instead of expecting liability, or any other singular action, to close the trust gap once and for all, medical institutions, AI companies, and regulators should collaborate.<sup>58</sup> A problem as complex as trust demands equally complex solutions.

### A. Certification Schemes

Certification schemes attest that a product has met certain safety or quality standards.<sup>59</sup> For example, consumers who care about sustainability can look for labels from organizations like the Rainforest Alliance or the Marine Stewardship Council.<sup>60</sup> Every label indicates that a government agency or private entity has audited the product to ensure compliance with the scheme's requirements.<sup>61</sup> In practice, certification schemes seem to help build trust.<sup>62</sup> Consumers, for example, tend to prefer products that carry certification labels.<sup>63</sup> Similarly, a certification scheme for AI could promote patients' trust.

---

56. Larry E. Ribstein, *Law v. Trust*, 81 B.U. L. REV. 553, 588 (2001).

57. Quinn et al., *supra* note 16, at 892 (cautioning that lack of clarity in liability could inhibit responsible use of medical AI, such as by promoting defensive medicine).

58. See WORLD ECON. F., EARNING TRUST FOR AI IN HEALTH: A COLLABORATIVE PATH FORWARD 8 (2025), [https://reports.weforum.org/docs/WEF\\_Earning\\_Trust\\_for\\_AI\\_in\\_Health\\_2025.pdf](https://reports.weforum.org/docs/WEF_Earning_Trust_for_AI_in_Health_2025.pdf) [<https://perma.cc/243X-W3R2>] (emphasizing the collective role of industry players and institutions in developing AI technologies that can earn patients' trust).

59. See *Product Certification Marks for Safety, Quality, and Performance*, INTERTEK (last visited Sept. 12, 2025), <https://www.intertek.com/product-certification-marks/> [<https://perma.cc/6E77-MFE8>].

60. See *Certifications to Look Out for when Purchasing Your Next Sustainable Product: What They Stand for*, GREEN HERMITAGE (May 24, 2023), <https://greenhermitage.com/en-us/blogs/blogs/certifications-to-look-out-for-when-purchasing-your-next-sustainable-product-what-they-stand-for/> [<https://perma.cc/8EH2-TYVY>].

61. CHRISTINE HAIGHT FARLEY, *Green Marks*, in RESEARCH HANDBOOK ON INTELLECTUAL PROPERTY AND CLIMATE CHANGE 399, 402 (Joshua D. Sarnoff ed., 2016); Jeanne C. Fromer, *The Unregulated Certification Market*, 69 STAN. L. REV. 121, 125–26 (2017).

62. Hall, *supra* note 3, at 501; Ribstein, *supra* note 56, at 586.

63. See Xiaoying Wang, *The Impact of Food Nutrition Labels on Consumer Behavior: A Cross-National Survey and Quantitative Analysis*, 1 INT'L J. PUB. HEALTH & MED. RSCH. 18, 19, 26 (2024).

A valid question here is how to articulate safety or reliability standards for AI, especially given its “black box” opacity and unexplainability. But evaluation methods are developing in the technical field. Autonomous vehicle companies constantly publish statistics to support their safety claims.<sup>64</sup> Accounting firms like Deloitte are racing to launch AI audit services, drawing from their experience in creating Environmental, Social, and Governance (“ESG”) metrics.<sup>65</sup> For AI in healthcare, a set of standards should someday be attainable, though I suspect that will not come from any one company, accounting firm, or regulatory agency. Instead, the industry will have to decide as a whole what safety really means, and what patients really are seeking.

Now, it might seem contradictory to suggest that certification schemes promote trust; after all, the point of certification schemes is to measure safety, reliability, or some other quality objectively, meaning they only attest to trustworthiness, which is not trust. Indeed, technically, certification schemes themselves cannot create trust. But they measure trustworthiness, which can be a reason to trust. Furthermore, unlike liability regimes, certification schemes do not carry the same negative associations that weaken trust, such as the image of something going wrong. That makes them a more constructive way to build trust in AI.

### B. “Just Culture”: Truth-Seeking, Not Fault-Blaming

When an aviation accident happens, the industry and authorities do not default to assigning liability. Instead, the focus is on figuring out what went wrong, and how to prevent it from happening again.<sup>66</sup> This so-called “just culture” assures pilots and crew members that candid disclosure does not bring on more punishment or liability, reducing their incentive to lie, in turn facilitating investigation.<sup>67</sup> Compensating victims is of course still important, but left to an industry fund or insurance.<sup>68</sup> By prioritizing prevention over punishment, over a few decades, air travel has quickly risen to be one of the safest means of

---

64. E.g., *Waymo Safety Impact*, WAYMO, <https://waymo.com/safety/impact/> (last visited Sept. 14, 2025) [<https://perma.cc/PMA3-3C5E>].

65. Ellesheva Kissin, *Big Four Firms Race to Develop Audits for AI Products*, FIN. TIMES (June 3, 2025), <https://www.ft.com/content/25b88580-1f89-491b-be91-ce0f2df95dfa> [<https://perma.cc/S7J5-SGGE>].

66. See, e.g., Paul Stephen Dempsey, *Independence of Aviation Safety Investigation Authorities: Keeping the Foxes From the Henhouse*, 75 J. AIR L. & COM. 223, 233 (2010) (discussing key provision in international aviation law that the purpose of accident investigations should be to prevent future accidents, not to assign blame or liability); Chloe A.S. Challinor, *Accident Investigators Are the Guardians of Public Safety: The Importance of Safeguarding the Independence of Air Accident Investigations as Illustrated by Recent Accidents*, 42 AIR & SPACE L. 43, 43–44 (2017); see generally Kyra Dempsey, *Why You’ve Never Been in a Plane Crash*, ASTERISK MAG. (Feb. 2024), <https://asteriskmag.com/issues/05/why-you-ve-never-been-in-a-plane-crash> [<https://perma.cc/GTQ3-5KG8>]; but see Mervyn E. Bennun & Gavin McKellar, *Flying Safely, The Prosecution of Pilots, and the ICAO Chicago Convention: Some Comparative Perspectives*, 74 J. AIR L. & COM. 737, 743–44 (2009) (contending that airlines’ “just culture” is not a “blame-free” culture).

67. *Id.*

68. See James Healy-Pratt & Owen Hanna, *How is Compensation Calculated After an Aviation Accident?*, KEYSTONE L. (Nov. 3, 2021), <https://www.keystonelaw.com/keynotes/how-is-compensation-calculated-after-an-aviation-accident> [<https://perma.cc/3Z9K-R4TM>].

transportation in history—much safer than driving, by comparison.<sup>69</sup> The ever-increasing aviation passenger volume attests to society's high level of trust.<sup>70</sup>

When healthcare AI inflicts harm, instead of attributing blame under liability regimes, perhaps a similar “just culture” would actually help make AI safer in the long run. But just as the aviation industry had to work together to establish “just culture” as its norm, the same collective effort is needed for healthcare AI.<sup>71</sup> Medical institutions, professionals who use AI, AI companies, and regulators must come together to decide what prevention over punishment looks like. That includes creating a compensation fund, establishing investigation principles, creating platforms to share lessons learned, and more. Bringing all these stakeholders together is no easy feat, but that effort is necessary to truly solve the problem of AI safety and earn patients' trust. A genuine, collective commitment to improve would itself be one of the strongest foundations for trust.

## VII. CONCLUSION

In the healthcare sector, trust is imperative but elusive. This is even more true once opaque, unpredictable AI enters the fray. Although liability is often society's response to risks and harm, it unfortunately does not create trust. It is sometimes good for incentivizing trustworthiness, but that is distinct from trust; indeed, it could even weaken trust. For a new technology to earn trust, finding a single scapegoat to bear liability is never going to suffice. Either formulating consistent, measurable standards for AI safety, or embracing an attitude of thorough investigation and accident prevention, the upshot is that trust takes an entire industry's hard work. Only when the healthcare sector takes trust seriously, and takes responsibility for earning patients' trust, can the virtues of trust continue to prevail in the AI era.

---

69. Christine Chung, *What You Should Know About Airplane Safety After Recent Crashes*, N.Y. TIMES (Feb. 18, 2025), <https://www.nytimes.com/2025/02/18/travel/plane-crashes-flight-safety.html> [https://perma.cc/F3AS-5ZDY]; Chris Isidore, *Flying Is Getting Scary. But Is It Still Safe?*, CNN (June 20, 2024, 2:02 PM EDT), <https://edition.cnn.com/2024/06/20/business/is-it-safe-to-fly-airplanes-boeing-max> [https://perma.cc/AUU6-QT2F].

70. Press release, International Air Transport Association, *Global Air Passenger Demand Reaches Record High in 2024* (Jan. 30, 2025), <https://www.iata.org/en/pressroom/2025-releases/2025-01-30-01/> [https://perma.cc/9YLN-8VJ3].

71. See Tony Ingesson, *Beyond Blame: What Investigations of Intelligence Failures Can Learn from Aviation Safety*, 35 INT'L J. INTELLIGENCE & COUNTERINTELLIGENCE 527, 530–31 (2022) (recounting the history of aviation investigations).