Michigan Technology Law Review

Article 4

2022

Degrees of Confidence as a Legal Tool to Assess AI System Liability

Joshua Song
University of Michigan Law School

DEGREES OF CONFIDENCE AS A LEGAL TOOL TO ASSESS AI SYSTEM LIABILITY

*Joshua Song**

ABSTRACT

AI systems have become increasingly integrated into our everyday lives, and harms caused by these systems have graduated from raising hypothetical ethical concerns to questions of actual legal liability. Civil liability schemes are generally designed to address harms caused by humans; thus, it may be tempting to analogize new types of harms caused by AI systems to familiar harms caused by humans in order to justify commandeering existing human-centered legal tools to assess AI liability. However, the analogy is inappropriate and misrepresents salient legal differences in how harms are committed by humans and AI systems. Thus, "as is often the case when analogical reasoning cannot justifiably stretch extant law to address novel legal questions raised by a new technology, new law is needed."[1]

First, I will discuss the legally salient difference between human and AI decision-making. Second, I will highlight two specific AI harms – autonomous vehicle product liability harms and predictive privacy harms – for which the analogy of human liability is insufficient. Finally, I will propose a new legal tool that may supplement the deficiencies in applying human liability schemes to AI harms.

* J.D. Candidate, 2023, University of Michigan Law School; B.E., 2017, University of Michigan.

1. Rebecca Crootof, *Autonomous Weapon Systems and the Limits of Analogy*, 9 HARV. NAT'L SEC. J. 51, 58–59 (2018).

TABLE OF CONTENTS

INTRODUCTION

Artificial intelligence systems are not monolithic in design, and employ different types of algorithms to produce different types of output variables.[2] Here, I will focus on machine learning algorithms which output categorical variables, a common type of AI system used in many real-world applications.[3] These machine learning systems generally function as follows: an algorithm is used to train a model on a large set of training data.[4] Once deployed, the model collects input data, performs a statistical analysis, and categorizes the data into classes with a certain degree of confidence.[5] The degree of confidence represents the estimated accuracy of the classification—in other words, the self-assessed reliability of the inference that the AI system makes.[6] For example, if an AI system classifies object X as an apple with a degree of confidence of 95%, this means the system has assessed a 95% probability that object X is an apple.

To demonstrate the entire lifecycle of a machine learning system from development to deployment, consider the computer vision systems in autonomous vehicles. Computer vision systems detect and classify obstacles per-

2. *See generally Output of AI Algorithms*, AI COLLECTIVE, https://www.aicollective.co/output-of-ai-algorithms (last visited Sept. 14, 2022).

3. *See generally* David Lehr & Paul Ohm, *Playing with the Data: What Legal Scholars Should Learn About Machine Learning*, 51 U.C. DAVIS L. REV. 653, 669–702 (2017).

4. *Id.* at 695–96.

5. *Id.* at 670.

6. IGI GLOBAL, *What is Degree of Confidence (DoC)*, https://www.igi-global.com/dictionary/degree-of-confidence-doc/42242.

ceived in the driving path of the vehicle.[7] In supervised learning, a large amount of training data is selected and classified; images are identified with appropriate labels defining what the image contains.[8] A model is then created through statistical and correlative analysis of the training data.[9] Once deployed, the model is able to take visual input from the vehicle's cameras and predict an appropriate classification for any objects contained in the input (e.g. 'car', 'human', 'animal', 'road').[10] Based on this classification, the autonomous vehicle makes a decision on what course of action to take (ignore, brake, speed up, etc.).[11] Like human decision-making, wrong decisions can lead to consequences. This includes civil harms for which the law needs appropriate tools to address.[12] However, human decision-making is fundamentally different from AI decision-making, which can render human-centered legal tools inappropriate when applied to AI decisions.

I. LEGALLY SALIENT DIFFERENCES BETWEEN HUMAN AND AI DECISION-MAKING

Many traditional human-centered legal tools used to assess civil liability require either an implicit or explicit statistical analysis. [13] Legal tools such as reasonable care may even prescribe a probabilistic standard.[14] However, facially strict quantitative standards are often applied with leniency when assessing human liability.[15] The reason for this leniency may be due to the nature and limitations of human decision-making and knowledge. The human understanding of probability can be infused with qualitative judgements and quantitatively irrational.[16] Humans may have a hard time accu-

7. *Introduction to Computer Vision for Self-Driving Cars*, THINK AUTONOMOUS (Jan. 24, 2018), https://www.thinkautonomous.ai/blog/?p=computer-vision-self-driving-cars-introduction.

8. *Id*.

9. *Id*.

10. *Id*.

11. *Id*.

12. *See generally Nearly 400 Car Crashes in 11 Months Involved Automated Tech, Companies Tell Regulators*, NAT'L PUB. RADIO (June 15, 2022), https://www.npr.org/2022/06/15/1105252793/nearly-400-car-crashes-in-11-months-involved-automated-tech-companies-tell-regul.

13. *See generally* Sarah Moss, PROBABILISTIC KNOWLEDGE 201–15 (2018).

14. *Id*.

15. *See* Sarah Moss, *Knowledge and Legal Proof*, 7 OXFORD STUD. IN EPISTEMOLOGY (forthcoming) (manuscript at 4). For example, "beyond a reasonable doubt" is a legal standard which is facially contingent upon a probabilistic degree of confidence threshold. However, in practice, the reasonable doubt threshold is hard to define, and in some jurisdictions, attempting to explain or define a strict reasonable doubt standard to the jury constitutes grounds for appeal.

16. *See* United States v. Hall, 854 F.2d 1036, 1044 (7th Cir. 1988) (Posner, J., concurring) ("Numerical estimates of probability are helpful in investments, gambling, scientific research, and many other activities but are not likely to be helpful in the setting of jury deliberations. No objective probability of a defendant's guilt can be estimated other than in the rare

rately defining quantitative factors, especially for in-the-moment decisions.[17] On the other hand, AI decision-making is purely quantitative: AI models set clearly-defined standards well before the actual moment of decision and AI systems base decisions on explicit probabilistic calculations made at speeds only achievable by computers.[18] Therefore, the implicit leniency provided in human-centered tools should not be applicable in the analogous legal tools for AI systems.

A. *Quantitative and Qualitative Probabilistic Analysis*

Human-centered legal tools are lenient toward quantitative standards because they reflect the qualitative nature of human decision-making. For example, the civil law 'preponderance of the evidence' standard incorporates a probabilistic analysis which asks the factfinder to determine whether it was more probable than not that a law was broken.[19] However, this legal tool which facially relies on quantitative statistical analysis also incorporates a hidden qualitative view of probability. The 'preponderance of the evidence' standard formally only requires a probability greater than 50% that a claim is true of a specific defendant in order to sustain a conviction.[20] In actuality, the 'preponderance of the evidence' standard does not function as a strict quantitative query.[21] Take for example the classic *Gatecrasher* hypothetical posed by Jonathan Cohen.[22] A proprietor of a rodeo sells 499 tickets to a rodeo show; on further inspection, the proprietor finds that there are actually 1000 people in attendance at the stadium and thus 501 are trespassing.[23] Without hearing any further evidence, a jury cannot convict a person randomly selected from the group of 1000 even though there is a 50.1% probability that the selected person is guilty of civil trespass; statistical evidence alone is insufficient to license a conviction verdict.[24] Consider also *Kaminsky v. Hertz Corp.*[25] There, the plaintiff was harmed in an accident caused by a yellow truck with a Hertz logo.[26] The defendant, Hertz Corp., owned 90% of vehicles fitting this description; thus, there was a 90% prob-

---

case that turns entirely on evidence whose accuracy can be rigorously expressed in statistical terms… It is one thing to tell jurors to set aside unreasonable doubts, another to tell them to determine whether the probability that the defendant is guilty is more than 75, or 95, or 99 percent.").

17. *Id*.
18. *See generally* Lehr & Ohm, *supra* note 3.
19. MOSS, *supra* note 13, at 202.
20. *Id*.
21. *Id*.
22. Jonathan L. Cohen, THE PROBABLE AND THE PROVABLE 74 (1977).
23. *Id*.
24. MOSS, *supra* note 13, at 207.
25. Kaminsky v. Hertz Corp., 288 N.W.2d 426 (1978).
26. *Id*. at 427.

ability that the individual tortfeasor was an employee of the defendant.[27] However, despite quantitative probabilistic evidence which far exceeded the demands of a preponderance standard, the court held that conviction on these grounds alone was not appropriate.[28] Not only does the preponderance standard reject conviction on statistical evidence alone, it actually allows for judgments free from any statistical support. A jury has free reign to determine when legal evidence has sufficiently satisfied the 'preponderance of the evidence' standard; a jury may convict a defendant even if they cannot provide any actual statistics or probability to support their analysis that the defendant was more likely than not to have committed the harm. This contradiction is no accident; the Supreme Court has affirmatively established that this facially quantitative legal standard is to be tempered by discretion and subjectivity.[29]

Thus, human-centered legal tools are often implemented with a *legal quantitative leniency*—legal tools are applied without strictly abiding by the quantitative standards that are facially required. This quantitative leniency is perhaps not unjustified. Human decision-making and knowledge are often disjointed from quantitative analysis in a way that would make it unreasonable or unjust to legally require strict adherence to quantitative standards. For instance, it would be untenable and unreasonable to require juries to produce a coherent and precise numerical justification for every conviction requiring a preponderance of the evidence.[30]

A more complex justification for legal quantitative leniency is that probabilistic legal tools are designed with recognition that humans temper probabilistic knowledge and decision-making with both reasonable and unreasonable qualitative judgments. For example, perhaps due process can be understood as an entirely reasonable insertion of qualitative analysis into probabilistic judgment. Due process requires that all quantitative evidence, no matter how conclusive, be filtered through a reasonable jury before it is

27. *Id*.

28. *Id*. at 427–29 (although the probabilistic evidence established a prima facie showing sufficient to prevent summary judgment that Hertz did not own the truck, the probabilistic evidence is merely sufficient for the case to proceed in a jury trial, and is not sufficient to justify the outright conclusion of Hertz's ownership as a matter of law).

29. *See* In re Winship, 397 U.S. 358, 370 (1970) ("a standard of proof represents an attempt to instruct the fact-finder concerning the degree of confidence our society thinks he should have in the correctness of factual conclusions for a particular type of adjudication. Although the phrases 'preponderance of the evidence' and 'proof beyond a reasonable doubt' are quantitatively imprecise, they do communicate to the finder of fact different notions concerning the degree of confidence he is expected to have in the correctness of his factual conclusions."); *see also* Allentown Mack Sales and Serv., Inc. v. Nat'l Lab. Rels. Bd., 522 U.S. 359, 373 (1998) ("it is conceivable that in certain categories of cases an adjudicating agency which purports to be applying a preponderance standard of proof might so consistently demand in fact more than a preponderance").

30. *See* United States v. Hall, *supra* note 16, at 1044.

accepted as legal fact.[31] Within the bounds of reasonableness, a jury is free to make any qualitative judgments about the quantitative evidence, even if such qualitative judgments are quantitatively erroneous.[32] For example, courts are often reluctant to convict on statistical evidence alone unless the statistical evidence is highly conclusive, such as with DNA profile matches.[33] Outside of such conclusive statistical evidence, due process invites jurors to insert their own qualitative judgment on the facts as a proxy for the qualitative judgment of society.[34]

The existence of forum shopping serves as an example of qualitative judgment within due process. Although the quantitative facts of a case remain the same, the qualitative judgments that society makes on the facts may differ by jurisdiction.[35] Where multiple courts have jurisdiction over the case, civil procedure allows litigants to seek a forum where the qualitative judgments are favorable towards their case.[36] The practice of forum shopping, coupled with the due process right to a jury trial, thus insert qualitative judgments into litigation, even those with quantitative standards of proof. [37]

Another justification for legal quantitative leniency is that humans do not process probability rationally and cannot be relied on to make accurate

---

31. In re Winship, *supra* note 29, at 364 ("Due process commands that no man shall lose his liberty unless the Government has borne the burden of convincing the factfinder of his guilt.") (internal quotations omitted).

32. *See id.* ("There is always in litigation a margin of error, representing error in fact finding, which both parties must take into account.").

33. MOSS, *supra* note 13, at 217–19.

34. In re Winship, *supra* note 29, at 370 (A juror's understanding of the applicable standard of proof may vary and lack definiteness, but it serves as a qualitative proxy for societies standard for conviction. "[T]he degree to which a factfinder is convinced that a given act actually occurred—can, of course, vary. In this regard, a standard of proof represents an attempt to instruct the fact-finder concerning the degree of confidence our society thinks he should have in the correctness of factual conclusions for a particular type of adjudication.").

35. For example, the Western District of Texas has emerged as a jurisdiction particularly favorable towards asserting patent rights. Compared to other jurisdictions, the Western District of Texas holds a broader definition of patentable subject matter and is much less likely to invalidate a patent on eligibility grounds. This has led to nearly a quarter of the nation's yearly patent infringement cases to be filed in the Western District of Texas. Paul R. Gugliuzza & J. Jonas Anderson, *How the West Became the East: The Patent Litigation Explosion in the Western District of Texas*, PATENTLY-O (Sept. 15, 2020), https://patentlyo.com/patent/2020/09/litigation-explosion-district.html.

36. *See Forum Shopping*, LEGAL INFORMATION INSTITUTE, https://www.law.cornell.edu/wex/forum_shopping

37. In some instances, legal theorists have asserted that due process may require qualitative judgments to completely override any factual or quantitative judgments. For example, Ronald Dworkin asserts that "a court ought not to convict, at least in some circumstances, even if it sustains the statutes and finds the facts as charged" where the defendant violates a morally doubtful law as a form of civil disobedience. Ronald Dworkin, *On Not Prosecuting Civil Disobedience*, THE N.Y. REV. OF BOOKS (June 6, 1968), https://www.nybooks.com/articles/1968/06/06/on-not-prosecuting-civil-disobedience/#:~:text=%E2%80%9C%5BIt%5D%20is%20of%20the,if%20a%20criminal%20conviction%20ensues.

quantitative judgments.[38] Humans are notorious for acting on inaccurate risk perceptions.[39] For example, humans unreasonably underestimate the risk of accidents when driving.[40] This unwarranted optimism is the result of qualitative factors such as perceived control, driving experience, and strong emotions.[41]

Therefore, human-centered quantitative legal tools are implemented leniently with an inherent understanding of the qualitative nature of human probabilistic analysis. However, AI decision-making and categorization is constructed solely from quantitative analysis without the influence of any qualitative factors.[42] AI developers may insert their own qualitative judgments into an AI system; for example, autonomous vehicle computer vision developers might set a lower degree of confidence threshold for classifying humans in order to initiate obstacle avoidance more frequently when the obstacle could be a human.[43] However, these qualitative judgments are made in the design of the system and must be translated to quantitative thresholds before the AI can act upon them.[44] Once deployed, the AI neither inserts its own qualitative judgements nor strays from the quantified thresholds in its design; its in-moment decisions are based solely on objective statistical analysis.[45] As such, once deployed, AI decision-making is purely quantitative and thus should not operate under the same quantitative leniency that human-centered legal tools employ.

B. *Temporal Proximity and Quantified Decision-Making*

Human-centered legal tools are lenient towards quantitative standards because it is difficult for humans to accurately make in-the-moment quantitative judgments. For example, in tort negligence claims, a fact-finder is asked to determine whether a decision conformed to the standard of care of a reasonable person.[46] This reasonableness determination is often quantified in a probabilistic test, the most notable being Judge Learned Hand's formula

38. *See* Morgan Housel, *Why We're Awful at Assessing Risk*, USA TODAY (Mar. 17, 2014), https://www.usatoday.com/story/money/personalfinance/2014/03/17/why-were-awful-at-assessing-risk/6530753.

39. *Id*.

40. *See* David M. DeJoy, *The Optimism Bias and Traffic Accident Risk Perception*, 21 ACCIDENT ANALYSIS & PREVENTION 333, 333–34 (1989).

41. *Id.* at 338-39; *see also Why Are Humans Bad at Calculating Risk?*, COGENCY (Feb. 23, 2018) ("People are less able to accurately assess probability when faced with either strong positive or negative emotions."), https://www.cogencyteam.com/news/2018/02/why-are-humans-bad-at-calculating-risk.

42. *See generally* Lehr & Ohm, *supra* note 3.

43. *Id.* at 674 (Autonomous vehicle system developers may set predictive goals for the AI system, such as minimizing human casualties).

44. *See, e.g., id.* (explaining the predictive goal of human casualties must be translated to an outcome variable on which input data is used to create a statistical model).

45. *Id*.

46. RESTATEMENT (THIRD) OF TORTS § 7 (Am. L. Inst. 2010).

for calculating negligence—the B<PL analysis.[47] Under this analysis, a risk bearing decision is unreasonable if the burden of precautions to avoid the risk (B) is less than the probability of the harm (P) multiplied by the gravity of the harm (L).[48] Important in this analysis is the decision maker's ability to accurately quantify the probability that a risk bearing decision will result in harm and to clearly define a PL threshold for when that risk becomes too great.[49] However, humans are often incapable of accurately assessing risk or defining risk thresholds.[50] This accuracy degrades further in split-second decisions.[51] These limitations in human decision-making constrain the reasonable application of human-centered legal tools. For example, it would be unreasonable to demand that a human driver conduct a B<PL analysis for every driving decision since the driver likely cannot accurately quantify the risk of each driving action and set reasonable risk thresholds in the short amount of time leading up to an automobile accident. Thus, quantitative leniency in human-centered legal tools may be justified by the temporal and cognitive limitations on humans' ability to process probability and other quantitative factors.

AI systems, however, do not face the same obstacles in probabilistic analysis. The bulk of an AI system's substantive decision-making process occurs long before an actual decision must be made.[52] AI developers design the algorithm, train the machine learning model, and set appropriate risk thresholds for decisions in far temporal proximity to the actual decision-making event.[53] Although the AI system must still make in-moment computations after deployment, such as interpreting and categorizing input, computers are not limited by computation speed in the same way that humans are and thus are not as restricted by the speed-accuracy tradeoff.[54] Furthermore, the in-moment computations should not be characterized as decisions *per se*. The design decisions and model training during the development of the AI constitute the substantive decision-making process of the system; these can be characterized as establishing a set of predetermined 'rules' for

47. *United States v. Carroll Towing Co.*, 159 F.2d 169, 173 (2d Cir. 1947).

48. *Id*.

49. *Id*.

50. *See generally* DeJoy, *supra* note 40.

51. *See* Richard P. Heitz, *The Speed-Accuracy Tradeoff: History, Physiology, Methodology, and Behavior*, FRONT. NEUROSCI., (June 11, 2014) (discussing the speed-accuracy tradeoff: the inverse correlation between decision speed and decision accuracy), https://www.frontiersin.org/articles/10.3389/fnins.2014.00150/full.

52. Lehr & Ohm, *supra* note 3, at 695-701.

53. *Id*.

54. *See* Heitz, *supra* note 51 (explaining the speed accuracy trade-off is a human limitation in neuro choice behavior).

the system to strictly follow after deployment in specific contexts without any further substantive decision-making or value judgments.[55]

AI systems are also designed to accurately quantify all factors considered in their decision-making. For example, when an autonomous vehicle makes driving decisions, its computer vision system categorizes any detected obstacles and the risks they present in degrees of confidence—a quantified probability reflecting how likely an obstacle belongs to a certain class of objects or how probable a driving decision will result in harm.[56] This quantitative degree of confidence of the AI system's categorization is compared against the pre-established degree of confidence threshold in order to facilitate the system's decision-making.[57] Although it is not guaranteed that AI systems accurately assess probability and other quantitative factors, the system's clearly quantified probability assessments that can be directly measured against the distinct quantified thresholds, providing a traceable and unambiguous roadmap for how the AI arrived at any decision.[58] Therefore, AI systems do not suffer from the same quantitative ambiguity in decision-making that plagues humans and warrants leniency in human-centered legal tools.

The above sections outline a few pertinent differences between AI decision-making and categorization and human decision-making and knowledge. Compared to humans, AI systems do not employ qualitative analysis, and have greater temporal proximity and ability to quantify variables in probabilistic decision-making. Although applying legal tools with quantitative leniency towards human decision-making and knowledge may be justified, the fundamental differences between humans and AI does not establish a sufficient legal analogy to justify a similar quantitative leniency when judging harms caused by AI systems.

II. INSUFFICIENT LEGAL ANALOGY AS APPLIED TO REAL HARMS

To illustrate the insufficiency of legal analogy between human and AI harms, I have chosen two specific AI harms to expand upon: predictive pri-

---

55. *See* Peter Railton, *Ethical Learning, Natural and Artificial, in* Ethics of Artificial Intelligence 45, 63 (S. Matthew Liao ed., 2020) (characterizing an autonomous vehicle system's approach to the trolley problem as a system following a rule in contrast to a human actor making a decision).

56. *See, e.g.*, Guy Tonye, *Machine Learning Confidence Scores — All You Need to Know as a Conversation Designer*, MEDIUM (Aug. 25, 2021) (Microsoft's conversational AI system compares a degree of confidence "confidence score" against pre-determined thresholds of 0.3 and 0.7.), https://medium.com/voice-tech-global/machine-learning-confidence-scores-all-you-need-to-know-as-a-conversation-designer-8babd39caae7.

57. THINK AUTONOMOUS, *supra* note 7.

58. *See generally* Lehr & Ohm, *supra* note 3; *see also* Robert Bielby, *Memory for Autonomous Vehicle Black-Box Recorders* (July 25, 2018), https://www.micron.com/about/blog/2018/july/memory-for-autonomous-vehicle-black-box-recorders.

vacy and autonomous vehicle product liability. Both of these AI harms have similar human counterparts, but nevertheless cannot be sufficiently addressed with human-centered legal tools. Many human-centered legal tools evaluate liability subject to either an implicit or explicit probability assessment, but as these two examples highlight, both implicit and explicit probability assessments in human-centered legal tools may produce inappropriate analogies to AI probability assessment.

In the first example, I argue that legal liability tools for human privacy harms require an implicit probability assessment on whether the confidence level in a given piece of information rises to the point of actual and actionable knowledge. Although AI predictive privacy harms may seem analogous to human privacy harms, the implicit knowledge threshold for humans cannot be equated to the AI knowledge and thus, the analogy is inappropriate. In the second example, legal liability tools for human negligence in automobile accidents require an explicit probability assessment to determine whether a risk taken by the driver was reasonable. Although the autonomous driving system may seem analogous to a human driver, AI systems and humans have fundamentally different approaches to probabilistic decision-making and thus, the analogy is inappropriate.

A. *Predictive Privacy Harms*

A primary value of AI systems is the ability to take in large amounts of data and make accurate predictions that humans otherwise could not accomplish.[59] However, when an AI system does its job too well, a new possible form of privacy harm arises. Predictive privacy harms ensue when AI systems have been able to take non-private data without any personally identifiable information and make probabilistic inferences on deeply personal and private information.[60] For example, in 2012, Target deployed a machine learning system that monitored its customers' in-store shopping habits to predict whether or not they were pregnant in order to mail pregnancy related coupons.[61] Target's AI system classified a girl as pregnant based on her shopping habits and consequently sent her pregnancy product coupons before either the girl or her father knew about the pregnancy.[62] Target's AI sys-

59. Kamales Lardi, *Understanding the Value of Artificial Intelligence Solutions in Your Business*, FORBES (Jan. 26, 2021, 9:00 AM) ("AI solutions enable deep insights from big data sources or rapidly combine large datasets to derive intelligent insights."), https://www.forbes.com/sites/forbesbusinesscouncil/2021/01/26/understanding-the-value-of-artificial-intelligence-solutions-in-your-business/?sh=5282c55c657f.

60. Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 93–95 (2014).

61. Charles Duhigg, *How Companies Learn Your Secrets*, THE NEW YORK TIMES (Feb. 16, 2012), https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=6&_r=1&hp.

62. *Id*.

tem has been described as committing a predictive privacy harm towards the girl by discerning her pregnancy status and using that information without her consent.[63] However, it is unclear when an AI system effectively 'knows' private information. When is the degree of confidence for an AI categorization sufficiently high enough to constitute legally knowing the predicted information? Human-centered privacy legal tools do not address the confidence level required for a person to legally 'know' private information since a human's assessment of whether they 'know' a certain piece of information incorporates both a qualitative and quantitative probability assessment of whether the information is true.[64] On the other hand, machine learning systems use a purely statistical process to classify categorical variables with a quantitatively defined degree of confidence.[65] AI systems do not employ any other qualitative judgment, thus it may be inappropriate to infer that AI systems 'know' information when they act on such information in the same way that one may infer that a human 'knows' information when they act on such information. For example, during the 2012 incident, a Target employee stated that an 87% degree of confidence that a woman was pregnant would warrant sending the woman pregnancy related coupons.[66] With such a low degree of confidence, it may be inappropriate to assert that Target had committed a predictive privacy harm; Target's AI system may not be legally accountable for knowing that a woman was pregnant when it only determined that there was an 87% probability that the woman was pregnant. Here, we already find dissonance when attempting to analogize AI predictive privacy harms to human-centered privacy harms.

In order to evaluate whether the analogy between human and AI privacy harms is tenable, first I will examine how human-centered legal tools address privacy protection and how such protection employs an implicit probability assessment. Then I will examine how AI degrees of confidence is an untenable analogy to human knowledge. In order to overcome the insufficient analogy, the legal tools used to assess predictive privacy harms must take into account the degree of confidence an AI system has in the accuracy of the information it discerns.

Human-centered legal tools for privacy protection in the United States operates under a sectoral approach; each industry sector is subject to its own discrete data privacy regime.[67] Although privacy regimes vary across sectors, the Fair Information Practice Principles ("FIPPs") provide guidelines

63. Crawford & Schultz, *supra* note 60, at 94–95.

64. Moss, *supra* note 15, at 21–24 (In human-centered legal tools, a justified credence does not constitute knowledge, since knowledge does not depend on quantitative credence alone.).

65. *See generally* Lehr & Ohm, *supra* note 3.

66. Duhigg, *supra* note 61.

67. Nicolas Terry, *Of Regulating Healthcare AI and Robots*, YALE J.L. & TECH. 133, 156 (2019).

for privacy regulations.[68] The FIPPs generally establish five core principles of privacy protection: (1) notice/awareness, (2) choice/consent, (3) access /participation; (4) integrity/security; and (5) enforcement/redress.[69] Although the FIPPS addresses data privacy by establishing rights for the individual, controls on information usage, data lifecycle, and enforcement, the FIPPs does not establish what constitutes the class of private information for which usage and access must be regulated.[70] The FIPPs assume access and knowledge of private information and provide guidelines on how to act given such knowledge.[71] For the purposes of this section, I am not concerned about the substantive content of information which makes it private. Instead, I aim to evaluate the level of knowledge of private information which constitutes sufficient access to trigger privacy concerns and regulation.

Human-centered privacy protection includes an implicit probability assessment on whether information constitutes knowledge. I propose that systemic privacy concerns are not triggered by the accuracy of an entity's knowledge of specific private information, but are instead triggered by the degree of confidence the entity holds in the accuracy of the private information. Here, the degree of confidence is established by the process used to discern or obtain the private information and reflects the probability that the discerned private information is accurate.[72] The following example is illustrative. Suppose there are two entities, A and B, which are trying to discern person C's 9-digit social security number. Entity A's process for doing so is simply to make a single random guess of the number. A's process has a one in a billion chance of yielding the correct social security number, and A has a 0.0000001% degree of confidence in the accuracy of their information. Here, we would not assert that A's random guess violates C's privacy even on the extremely unlikely chance that A has guessed the correct number. We might say that A has violated C's privacy once A has confirmed that his guess was correct, but in this instance, A's information discerning process has changed; once A verifies the private information, the degree of confidence is much closer to 100%. If A acts on his guess without verification, A is operating with a 0.0000001% degree of confidence and thus does not have sufficient access or knowledge of private information to violate C's privacy.

On the other hand, imagine entity B has a process for discerning social security numbers with a 99.9% accuracy rate. Here, B has a 99.9% degree

68. Crawford & Schultz, *supra* note 60, at 107.

69. *Id*.

70. *See generally id*.

71. *See generally* FED. TRADE COMM'N, PRIVACY ONLINE: A REPORT TO CONGRESS 7–11 (1998), https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf.

72. Jason Brownlee, *Confidence Intervals for Machine Learning*, MACHINE LEARNING MASTERY (last updated Aug. 8, 2019) (explaining the function of degree of confidence within AI classification), https://machinelearningmastery.com/confidence-intervals-for-machine-learning.

of confidence in the accuracy of the number, and we can assert that B has violated C's privacy if B uses the number without consent regardless of the accuracy of the number. Indeed, it is intuitive to assert a violation of privacy even before any verification for whether the number is accurate occurs.

When faced with human privacy harms, it is often reasonable to assume knowledge without quantified analysis of the degree of confidence. Even though human knowledge is also inherently probabilistic, and one cannot truly be 100% confident in the accuracy of their knowledge, a sufficient probability is enough to justify actual knowledge.[73] However, before asserting a human privacy harm, a plaintiff does not have to justify that the defendant's degree of confidence in the accuracy of the private information was sufficiently high enough to constitute actual knowledge.[74] This leniency and assumption of knowledge may be due to the fact that it is difficult for humans to accurately quantify the degree of confidence they have in any piece of information and the perceived degree of confidence may be influenced by qualitative factors such as emotions or societal custom.[75] A legal assertion of human knowledge does not depend upon the degree of confidence in the knowledge; additionally, a high degree of confidence alone may not justify an assertion of knowledge.[76]

In contrast, a classification by an AI system cannot be assumed to legally constitute knowledge even if this classification is used in further decision-making by the system. An AI system which acts on a statistical inference is not conclusive proof that the inference constitutes knowledge discerned by the system. For instance, if the Target AI system was merely designed to provide coupon suggestions, the system does not have to rely on a high degree of confidence before acting. If, in an effort to cast a wide marketing net, Target may set the degree of confidence threshold to twenty percent and send a coupon to any woman that has a twenty percent chance of being pregnant. Such low confidence in the accuracy of the pregnancy classification is more akin to a mere hypothesis than legally actionable knowledge. Although there is no knowledge threshold in human privacy regulation, AI privacy regulation may use discretely defined degrees of confidence as a threshold to provide greater certainty and clarification. Regula-

73. MOSS, *supra* note 13, at 109–11.

74. *OCR Announces 4 Financial Penalties to Resolve HIPAA Violations*, HIPAA JOURNAL (Mar. 29, 2022) (The Office for Civil Rights (OCR) penalized Dr. U. Phillip Igbinadolor for violating the HIPAA Privacy Rule when Igbinadolor's office responded to a negative google review with the patient's name and symptoms. The patient left the review under a pseudonym, but OCR did not have to justify that Igbinadolor's degree of confidence was sufficiently high that the review correlated to a specific patient in order to assert that Igbinadolor violated the patient's privacy.), https://www.hipaajournal.com/ocr-announces-4-financial-penalties-to-resolve-hipaa-violations.

75. *See* COGENCY, *supra* note 41.

76. *See* Moss, *supra* note 15, at 22-24 ("[J]ustified credence does not constitute knowledge.").

tors may such thresholds to decrease ambiguity in what constitutes a predictive privacy harm and to delineate between actionable and unactionable AI inferences and classifications.

B. *Autonomous Vehicle Product Liability Harms*

Motor vehicles with autonomous driving capabilities which once existed only on the pages of sci-fi novels now exist in rapidly increasing numbers on our roads and highways.[77] The rise of autonomous vehicles unfortunately also came with a rise in autonomous vehicle crashes and harms.[78] Autonomous vehicle harms cannot be sufficiently addressed with human-centered legal tools. Human-centered auto accident liability legal schemes are well established, and thus it is tempting to simply apply these schemes to autonomous vehicles.[79] Indeed, the explicit negligence and reasonable care inquiries in auto liability evaluate a human actor's probabilistic risk analysis in a form that seemingly resembles the probability analysis undertaken by an autonomous driving system. However, salient differences in human and AI probabilistic analysis render the analogy inappropriate. In this section, first I will evaluate the existing human-centered auto liability schemes and the explicit probability analysis that they present. Then, for each scheme, I will analyze how autonomous vehicle liability may fit into the general existing structure of auto liability, but nevertheless cannot completely adopt the entirety of the liability system as the AI probability analysis and decision-making is substantively different from their human counterparts. Generally, existing liability schemes involve significant qualitative discretion by the factfinder and are lenient in their purported quantitative standards.[80] Autonomous vehicles are designed with clear quantitative standards and decision-making processes—the machine learning computer vision system clearly defines the degree of confidence in the classification of each object it detects and acts according to predetermined probability thresholds set by the developers.[81] Therefore, autonomous vehicle liability schemes do not need to incorporate the quantitative leniency present in human-centered

77. *See Autonomous/Driverless Car Market – Growth, Trends, COVID-19 Impact, and Forecast (2021-2026)*, REPORT LINKER (June 2021) ("[A]utonomous cars witnessed a rise in sales, as in 2020, around 11.2 million cars were sold with level 2 features, which is an increase of 78% from 2019.").

78. *See generally* NAT'L PUB. RADIO.

79. *See* Nora Freeman Engstrom, *When Cars Crash: The Automobile's Tort Law Legacy*, 53 WAKE FOREST L. REV. 293, 295 (2018) (Automobile accident lawsuits account for "half of all trials, nearly two-thirds of all injury claims, and three quarters of all damage payouts.").

80. *See Kaminsky*, *supra* note 25, at 427-29; *see also* Engstrom, *supra* note 79, at 316-20 (Discussing *Goodman* and *Pokora*, and the progress of automobile tort liability evolving away from strict negligence rules and towards negligence standards to be applied by juries.).

81. *See* THINK AUTONOMOUS, *supra* note 7.

schemes and can prescribe strict probability standards for negligence and reasonable care.

Existing auto accident liability schemes generally exist in two categories.[82] Driver-centered schemes such as reckless or negligent driving focus on the harm committed by the driver, and manufacturer centered schemes such as product liability focus on the harm committed by the manufacturer.[83] When considering accidents involving autonomous vehicles, it is tempting to repurpose the existing liability schemes. While the general framework of a driver-centered or manufacturer-centered scheme may be useful to frame autonomous vehicle liability, the human-centered legal tools used in these schemes to assess reasonable decision-making are inappropriate as applied to autonomous vehicles.

While uncommon today, the possibility exists for liability to be assessed in autonomous vehicle accidents through driver-centered schemes such as reckless or negligent driving. The autonomous driving system itself may qualify as a 'driver' depending on its SAE level of autonomy.[84] Thus, the 'driver' may be liable for reckless or negligent driving, and indeed, courts have already seen complaints alleging liability for negligent driving by autonomous driving systems.[85]

Reckless driving is generally defined by most states as driving with a willful or wanton disregard for the safety of persons or property.[86] Thus, human-centered reckless driving requires a reckless, malicious, or willful mental state.[87] While determining the mental state of an AI system is already nonsensical, it is even more inconceivable to try to determine if an AI system had any malicious or reckless motivations. Therefore, human centered legal tools for reckless driving are untenable as applied to autonomous driving systems.

---

82. Engstrom, *supra* note 79, at 308 n.90 ("[O]ne injured motorist's claim can give rise to an auto claim against the negligent motorist as well as a product liability design defect claim on the theory that the car was insufficiently crashworthy.").

83. *Id*.

84. Letter from Paul A. Hemmersbaugh, Chief Counsel, National Highway Traffic Safety Administration, to Chris Urmson, Self-Driving Car Project Director, Google, Inc. (Feb. 4, 2016) (In a 2016 interpretation letter to Google regarding Google's Waymo self-driving vehicle project, NHTSA determined that for the purposes of Federal Motor Vehicle Safety Standards, the Google AI driving system was to be classified as the 'driver' of the vehicle).

85. *See* Complaint at 15-16, Nilsson v. General Motors LLC, No. 3:18-cv-00471 (N.D. Cal. Jan. 22, 2018), 2018 WL 514625 ("Defendant owed Plaintiff a duty of care in having its Self-Driving Vehicle operate in a manner in which it obeys the traffic laws and regulations. Defendant breached that duty in that its Self-Driving Vehicle drove in such a negligent manner that it veered into an adjacent lane of traffic without regard for a passing motorist, striking Mr. Nilsson and knocking him to the ground.").

86. *See* Michigan Vehicle Code, MICH. COMP. LAWS § 257.626 (2011) ("a person who operates a vehicle upon a highway… in willful or wanton disregard for the safety of persons or property is guilty of a misdemeanor."); Reckless Driving, CAL. VEH. CODE § 23103(a) (West 2011) ("A person who drives a vehicle upon a highway in willful or wanton disregard for the safety of persons or property is guilty of reckless driving.").

87. *See Wanton*, BLACK'S LAW DICTIONARY (11th ed. 2019).

Negligent driving is generally defined as driving in a careless or negligent manner likely to endanger any person or property.[88] Although negligence may sometimes be inferred through *res ipsa loquitur*, the legal principle that the occurrence of a harm implies negligence, automobile accidents generally do not apply the *res ipsa* doctrine and instead inquire whether the driver failed to exercise a standard of care that a reasonably prudent person would have exercised in the same situation.[89] Reasonable care as a human centered legal tool is inappropriate when applied to autonomous driving systems. Autonomous driving systems should not be held to the standard of driving of a reasonably prudent person. As discussed above, even the most prudent humans are subject to inaccuracies and limitations in risk assessment and split-second decision-making.[90] As autonomous driving systems are not subject to the same limitations that humans are, inquiring whether an autonomous vehicle merely acted as a reasonable human would is too lenient and would perhaps foreclose any finding of negligence. Therefore, the analogy between human driving and AI system driving is inappropriate and insufficient to justify using human-centered recklessness or negligence liability schemes to assess autonomous vehicle driver centered liability.

It is important to note that, as applied, autonomous vehicle reckless or negligent driving liability schemes may function very similarly to product liability schemes. Even if the autonomous driving system is classified as the 'driver' in an accident, the autonomous driving system is not only just the driver of the specific vehicle in the crash. The same autonomous driving system is the 'driver' of all vehicles which employ the same version of the driving program. Convicting one particular instantiation of the autonomous driving system of reckless or negligent driving would likely imply that all autonomous driving systems operating under the same programming would be liable if an accident occurred under similar circumstances. Additionally, the autonomous driving system itself cannot shoulder any liability even if it is classified as the 'driver.' Thus, the burden for any damages due to reckless or negligent driving would likely fall to the autonomous driving system manufacturer or the original equipment manufacturer. Perhaps either for these reasons, or because of the unworkable comparison between a reasonable human driver and an autonomous driving system, the focus for assessing liability in autonomous vehicle accidents is generally on manufacturer centered product liability schemes.

88. *See* Michigan Vehicle Code, MICH. COMP. LAWS § 257.626b (1979) ("A person who operates a vehicle… in a careless or negligent manner likely to endanger any person or property… is responsible for a civil infraction.").

89. *See Negligence*, BLACK'S LAW DICTIONARY (11th ed. 2019); Judicial Council of California Civil Jury Instructions No. 401 (2021) ("Negligence is the failure to use reasonable care to prevent harm to oneself or to others.").

90. *See* DeJoy, *supra* note 40; *see* Heitz, *supra* note 51.

There are three categories of product liability: manufacturing defects, design defects, and defects in warnings or instructions.[91] Most pertinent to autonomous vehicle accidents are design defects, which occurs when "the foreseeable risks of harm posed by the product could have been reduced or avoided by the adoption of a reasonable alternative design… and the omission of the alternative design renders the product not *reasonably* safe."[92] For instance, when an autonomous vehicle is involved in an accident, if the design of the autonomous driving system is unreasonable, then the manufacturer may be subject to product liability.[93] In the autonomous vehicle context, the design may be unreasonable if the computer vision system unreasonably classifies obstacles, or if the driving system makes unreasonable driving decisions.[94] However, a general difficulty of assessing product design defects is that liability is only assigned if there is a reasonable alternative design that would avoid the harm. With new technologies that are still in development, there may not be many alternative designs; there are very limited alternative autonomous driving systems that are operational and deployed in the market.[95] Even if an alternative design exists, it would be difficult for a factfinder to determine whether the alternative design is reasonable. It would also be difficult to assess whether an alternate autonomous driving system design would reduce the chance of harm in a similar context.

Regardless of the difficulty, design defect product liability schemes ask whether an autonomous driving system design is reasonably safe or defective.[96] To assess whether a vehicle design is reasonably safe, courts apply either the consumer expectation test or the risk-benefit test.[97] The consumer expectation test asks whether the vehicle performs as safely as an ordinary consumer would reasonably expect it to perform.[98] The risk-benefit test asks whether the benefits of the vehicle design reasonably outweigh the risks of

---

91. RESTATEMENT (THIRD) OF TORTS: PRODUCTS LIABILITY § 2(b) (Am. L. Inst. 2010).

92. *Id*. (emphasis added).

93. Stephen S. Wu, *Product Liability Issues in the U.S. and Associated Risk Management, in* AUTONOMOUS DRIVING 559-60 (Markus Maurer et al. eds., 2016).

94. *Id*. at 563-64.

95. GRAND VIEW RESEARCH, AUTONOMOUS VEHICLES MARKET SIZE, SHARE, & TRENDS ANALYSIS REPORT BY APPLICATION (TRANSPORTATION, DEFENSE), BY REGION (NORTH AMERICA, EUROPE, ASIA PACIFIC, SOUTH AMERICA, MEA) AND SEGMENT FORECASTS 2022 – 2030 (2022) (Reporting that there are only 13 key companies which are operating in the global autonomous vehicle market.), https://www.grandviewresearch.com/industry-analysis/autonomous-vehicles-market#:~:text=The%20global%20autonomous%20vehicle%20market,103.4%20thousand%20units%20in%202022.95.

96. WU, *supra* note 93, at 560.

97. Barker v. Lull Engineering Co., 20 Cal. 3d 413, 415 (1978).

98. *Id. See also* Saller v. Crown Cork & Seal Co., 115 Cal. Rptr. 3d 151, 160 (Cal. Ct. App. 2010); Pannu v. Land Rover North America, Inc., 120 Cal. Rptr. 3d 605, 616 (Cal. Ct. App. 2011).

danger in the design.[99] Both of these tests are questions of fact which may become extremely difficult to answer when the product in question is an AI system.

The consumer expectation test incorporates "standardless references to the expectations of an ordinary consumer" which provides an ambiguous guide for liability, especially with new technologies where consumers have not formed cohesive expectations.[100] With an autonomous vehicle, the question is whether the computer vision system classified obstacles with the accuracy that an ordinary consumer would expect or whether the autonomous driving system made driving decisions based on the degree of confidence of classified obstacles that an ordinary consumer would expect.[101] There would be an undesirable chilling effect on autonomous vehicle development if the expectation was 100% accuracy or strict liability for any accidents.[102] However, since the technology is still new and developing, there is likely no established consumer expectation for permissible computer vision accuracy or autonomous vehicle driving decisions. The factfinder would have no comparison to draw upon beside the expectations of how a human driver would act in a similar situation. As discussed above, the comparison between a reasonable human and an autonomous driving system is an untenable standard for liability. Furthermore, here the comparison would be between the expectations of an individual human actor and the expectations of a product. Therefore, the consumer expectation test is inappropriate for autonomous vehicle product liability.

The risk-benefit test purports to rely more on quantitative utilitarian calculations; however, there is no clear standard for weighing the benefits and risks for new technology such as autonomous vehicles, and a strict application of the risk-benefit test may be too lenient towards autonomous vehicles. If the risks and benefits of the design are evaluated when considering the autonomous driving system as a whole, the risk-benefit test may be too lenient and frequently weigh in favor of the autonomous driving system. Autonomous vehicles and their development provide great benefit to the general public, and the risks are relatively minimal; in fact, autonomous vehicles may make safer driving decisions than humans do.[103] If the risks and benefits of the design are evaluated by considering individual decision-

99. *Saller*, *supra* note 98, at 161.

100. *Id*.

101. Wu, *supra* note 93, at 563-64 (Noting possible AV design defects to include inaccurate pattern and obstacle recognition and failure to perform safe ordinary driving maneuvers).

102. *See also id*. at 553 ("The threat of massive product liability litigation involving AVs is widely perceived as one of the chief obstacles to AV development and sales, if not the number one threat.").

103. *See* NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., *Automated Vehicles for Safety* (last visited May 12, 2022) (discussing heightened safety as one of the biggest benefits of autonomous driving systems), https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety.

making units within the larger autonomous driving system design, the nature of emerging AI technology may make risks and benefits too hard to discern. The comparison of risks and benefits may then entail a policy question that is inappropriate for a jury to decide. For instance, an autonomous driving system design needs to set a degree of confidence threshold for obstacles classified as humans.[104] Above this threshold, the autonomous driving system will recognize that there is a human in the driving path of the vehicle and act to avoid the obstacle; below this threshold, the autonomous driving system will not recognize a human in the driving path and will decide not to use any evasive maneuvers.[105] If the system designers set the threshold to be 0.5%, this means that the autonomous driving system will avoid any objects that it believes has at least a 0.5% probability of being a human and will not treat any object with less than 0.5% probability of being a human as a human.[106] Under a risk-benefit test, the question becomes, when employing a computer vision system to classify humans, what is the confidence rate that will shield autonomous vehicles from liability when they hit a human? This question may involve too great of a policy consideration for an individual jury to decide; AI systems which rely on a probabilistic degree of confidence determination in order to make any decision will consistently make the same decision given the same degree of confidence and the same programming across all instances and all products which deploy the same AI system. Making a risk-benefit reasonableness judgement on one instance of a degree of confidence threshold would set a liability standard for all AI systems employing the same programming. Thus, risk-benefit tests may not be appropriate to assess the product liability for autonomous vehicles. Since the consumer expectation and risk-benefit tests involve inappropriate analogies between human and AI driving decisions or require nebulous determinations by the factfinder, the current product liability schemes are insufficient to assess autonomous vehicle liability.

III. A PROPOSAL FOR A DEGREE OF CONFIDENCE STANDARD

Although the direct application of human-centered legal tools to AI systems is misguided, AI degrees of confidence may be used as a legal tool to supplement the insufficiencies and allow AI liability to remain within the existing human-centered legal framework. My proposal comes in two parts:

---

104. *See generally* THINK AUTONOMOUS, *supra* note 7 (discussing computer vision obstacle recognition).

105. *See* Sigil Wen, *Object Detection with YOLO | Bringing Vision to Self-Driving Cars*, MEDIUM (Nov. 2, 2019) (images detected by an autonomous vehicle are categorized as different classes of objects based on whether the image meets a certain confidence score threshold for the object class.), https://towardsdatascience.com/object-detection-with-yolo-bringing-vision-to-self-driving-cars-980295226830.

106. *Id*.

(1) standards and regulations for acceptable degrees of confidence in specific contexts, and (2) AI 'expert witnesses' as a new legal tool for assessing liability. I will discuss both steps in turn in the following sections.

A. *Standards and Regulations for Acceptable Degrees of Confidence*

Standards and regulations already exist for many contexts in which AI systems are deployed.[107] However, those standards do not set bright line thresholds governing reasonable or acceptable degrees of confidence of an AI system.[108] FIPPs provide guidelines for privacy practices, but do not set a degree of confidence that would legally justify an AI system 'knowing' private information.[109] The Federal Motor Vehicle Safety Standards (FMVSS) promulgated by the National Highway Traffic Safety Administration (NHTSA) set performance standards for autonomous vehicles,[110] but do not establish degrees of confidence for acceptable decision-making.[111] Although these shortcomings render the standards and regulations ineffective for regulating AI liability, these deficiencies may be overcome by specifically addressing AI degrees of confidence in order to direct specific regulation towards the aspect of AI systems that are incompatible with human-centered legal tools.

I propose that administrative agencies and other regulatory bodies establish industry and context specific degree of confidence thresholds which will serve as clearly enumerated liability standards. For example, NHTSA would incorporate a degree of confidence schedule into the FMVSS to govern autonomous driving system liability. The schedule would set degree of confidence thresholds for each object class used in computer vision systems for classifying obstacles. To avoid liability, the autonomous driving system must make appropriate driving decisions to avoid the obstacle when it classifies the obstacle at a degree of confidence higher than the FMVSS threshold. The autonomous driving system would not be liable for failure to avoid an obstacle if it classified the object at a degree of confidence lower than the FMVSS threshold, even if the failure to avoid the obstacle resulted in a collision.

---

107. *See generally Legislation Related to Artificial Intelligence*, NAT'L CONF. OF STATE LEGIS. (Jan. 5, 2022), https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx.

108. *Id*.

109. *See generally* Crawford & Schultz, *supra* note 60, at 107-08.

110. *See generally Regulations*, NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., https://www.nhtsa.gov/laws-regulations/fmvss (last visited July 9, 2022); NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., FMVSS CONSIDERATIONS FOR VEHICLES WITH AUTOMATED DRIVING SYS.: VOLUME 1 (2020) [hereinafter AUTONOMOUS VEHICLE FMVSS], https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/ads-dv_fmvss_vol1-042320-v8-tag.pdf.

111. *See generally* AUTONOMOUS VEHICLE FMVSS, *supra* note 110 (The FMVSS does not discuss any standards for AI system design regarding degrees of confidence or confidence intervals).

In regard to privacy, regulation varies across sectors, but sector-specific degree of confidence standards can establish clear thresholds for when AI inferences become actual knowledge subject to privacy regulation. For example, in the health sector, an AI privacy concern is the re-identification of de-identified health data.[112] HIPAA regulated entities do not need to comply with the HIPAA privacy rule for health information that has been deidentified by removing 18 personal identifiers.[113] However, AI can be used to reidentify data; machine learning systems can be used to make inferences on the 18 removed personal identifiers, thus potentially eliminating the privacy protection of removing the identifiers.[114] To combat the privacy harms caused by AI reidentification, the Department of Health and Human Services (HHS) could establish a degree of confidence regulation for reidentified data. If deidentified personal data was reidentified at a degree of confidence above the HHS threshold, the reidentified data and personal identifiers are to be treated as private data subject to the HIPAA privacy rules, even if the reidentified data is actually inaccurate.[115]

These degree of confidence standards must be set by regulatory bodies instead of relying on the common law process of establishing reasonable degrees of confidence through piecemeal adjudication. The common law process is backward-facing and requires adjudication on a harm that has already occurred before standards can be set.[116] AI is a developing field where industry customs and consumer expectations are not yet established. Standards that can guide the path of AI development must be set in place before harms from unreasonable AI systems can occur. In an emergent field, this also provides the chance for law to drive the evolution of AI expectations and safeguards instead of the AI industry driving the evolution of AI law.

An immediate criticism of this proposal is that specific degree of confidence standards are hard to set and require detailed regulation in all sectors. This may be true. However, AI technology is rapidly being incorporated in-

112. *See* Nicholson Price, *Problematic Interactions Between AI and Health Privacy*, 2021 UTAH L. REV. 925, 926-27 (2021).

113. *Id*. at 926.

114. *Id*. at 926–27.

115. HHS currently regulates health data reidentification through HIPAA § 164.514(c) which states that if a covered entity successfully reidentifies health data, the data once again becomes private health information protected by the HIPAA privacy rule. However, under this statute, reidentification means actual success in discerning the identity of the health data instead of a sufficiently high degree of confidence in the identity of the health data. *See Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, U.S. DEPT. OF HEALTH AND HUMAN SERVS. (May 31, 2022), https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#:~:text=Under%20this%20standard%2C%20health%20information,used%20to%20identify%20an%20individual.

116. *See* Lujan v. Defs. of Wildlife, 504 U.S. 555, 560 (1992) (Standing in federal court requires that the plaintiff has suffered an injury in fact that is concrete and particularized, and actual or imminent, not conjectural or hypothetical).

to all sectors: commercial and non-commercial, private and public.[117] It is not inconceivable to think that AI will soon become a component of most aspects of our lives.[118] AI is no longer a legal side project for which general rules and vague guidelines will suffice. With such impact and prevalence, AI demands industry and context specific regulation across all sectors. There needs to be specific degree of confidence standards delineating, for example, between when AI inferences on pregnancy status constitute knowledge for the purposes of sending marketing coupons and when they constitute knowledge for the purposes of generating medical information. Additionally, administrative agencies and regulatory bodies are already familiar with the process of establishing detailed and specific standards in their respective industries. NHTSA has promulgated and continues to promulgate performance standards governing every minute detail of motor vehicle safety.[119] Degree of confidence standards for autonomous vehicle safety would be well within NHTSA's job description and regulatory power.

B. *AI Expert Witnesses*

Once industry specific standards for degrees of confidence are established, when a harm occurs, liability can be assessed by determining whether the defendant AI system complied with the applicable degree of confidence. The defendant would be liable if the inference or decision which caused the harm did not comply with the regulation governing the pertinent degree of confidence assessed by the AI system. This determination should be made, not only by evaluating the degree of confidence assessed by the AI system itself, but also by employing the aid of AI expert witnesses. AI expert witnesses are other third-party AI systems which operate in the same industry and within the same context as the AI system which caused the original harm. The AI expert witnesses are provided with the same input data, and their 'testimony' is the corresponding output degree of confidence for the pertinent AI inference or decision in question. If the defendant's assessed degree of confidence is reasonably similar to the AI expert witnesses' assessed degree of confidence, then the defendant's own degree of confidence is used to determine whether the defendant AI system complied with the pertinent regulation. However, if the defendant's assessed degree of

117. *See generally* Ron Schmelzer, *The Increasing Expansion of AI in Business and Government – Insights From AI World*, FORBES (Mar. 22, 2019), https://www.forbes.com/sites/cognitiveworld/2019/03/22/the-increasing-expansion-of-ai-in-business-and-government-insights-from-ai-world/?sh=3b8a49fd5def.

118. *See generally* Allen Victor, *10 Uses of Artificial Intelligence in Day to Day Life*, DAFFODIL (July 24, 2021) (Outlining common facets of our lives which have already incorporated AI applications), https://insights.daffodilsw.com/blog/10-uses-of-artificial-intelligence-in-day-to-day-life.

119. *See generally* NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., *supra* note 110; *see generally* AUTONOMOUS VEHICLE FMVSS, *supra* note 110.

confidence is unreasonably lower than that of the AI expert witnesses, an aggregation of the AI expert witnesses' degree of confidence for the specific inference is used to assess whether the defendant's inference complied with the relevant regulation. This prevents AI developers from attempting to escape legal liability by purposely reducing the accuracy and efficacy of their AI system.

Consider the following hypothetical. Assume an FMVSS standard provides that an autonomous vehicle must break if it is confronted with an obstacle that is categorized as a human with 0.5% degree of confidence or higher. If a crash occurs where an autonomous vehicle hits a human because the vehicle did not brake, liability would be assessed through the conjunction of two tests. (1) Does the vehicle's event data recorder show that the autonomous driving system assessed the obstacle to be a human with a degree of confidence greater than 0.5% and did not brake? (2) What is the applicable degree of confidence concluded by other similar autonomous driving systems on the market when given the same input data from the crash? Even if the defendant autonomous driving system had determined that the degree of confidence was lower than 0.5%, if the AI expert witnesses all concur that the degree of confidence was higher than 0.5%, then the defendant system would still be liable for violating the regulation.

In comparison to human expert witnesses, AI expert witnesses may be cheaper and faster. The AI expert witnesses are models which are already trained in the appropriate context and merely need the pertinent input data in order to produce a degree of confidence output. Additionally, AI expert witnesses are less subject to bad faith prejudice or unfair partiality towards one party over the other.[120] The AI expert witnesses would be taken directly from AI systems currently operating in the market and cannot change their programming to favor one party over the other. Most importantly, establishing concrete degree of confidence regulations and employing AI expert witnesses to audit compliance avoids any inappropriate legal analogy between humans and AI systems; AI knowledge and decision-making will no longer be judged by legal standards made for human knowledge and decision-making. Instead, this two-part proposal establishes probabilistic standards and assesses compliance through quantitative conformity to acceptable AI decision-making.

C. *Potential Criticisms*

This two-part proposal for establishing degree of confidence regulations and assessing AI liability through AI expert witnesses is likely ambitious and subject to multiple criticisms, but these criticisms may actually high-

120. Jason M. Chin et al., *The Biases of Experts: An Empirical Analysis of Expert Witness Challenges*, 42 Manitoba L.J. 21, 24-28 (2019) (Discussing the various types of bias which have unjustly colored human expert testimony and expert evidence).

light benefits of the proposal instead of limitations. One potential criticism is that the AI expert witness scheme would generate an anti-innovation chilling effect. If AI liability is determined by the accuracy of AI systems on the market, and generally a lower degree of confidence means less liability, then the AI expert witness liability scheme would disincentivize developers from making progress towards increasing accuracy and efficacy of the AI systems. In other words, an inaccurate AI system with poor decision-making could escape liability if all the AI expert witnesses were also inaccurate and poor at decision-making. Therefore, developers may be disincentivized from improving their AI system in order to avoid raising the bar for liability.

However, this anti-innovation chilling is unlikely to occur in a free market. Consider two companies, company A and company B which produce similar AI systems and operate in the same industry. If A's AI system is more accurate than B's system, A is less likely to be liable for harm when B's system is serving as an expert witness. B's assessed degree of confidence would generally be lower than A's assessed degree of confidence. Therefore, as an AI expert witness, B's degree of confidence would provide a lenient threshold that A is likely to meet. Additionally, B is more likely to be liable for harm when A's system is serving as an expert witness. A's assessed degree of confidence would be higher than B's assessed degree of confidence. Therefore, there would likely be many situations where B's assessed degree of confidence would indicate that B complies with regulation, but A's assessed degree of confidence of the input data would indicate that B does not comply with regulation. Thus, companies would gain a market edge by increasing the accuracy of the AI systems they develop in order to decrease the likelihood of liability for their own product while increasing the likelihood of liability for their competitors' AI systems.

Another potential criticism of the AI expert witness proposal is that this liability scheme would require widespread collection and transparency of input data. In order for AI expert witnesses to assess the pertinent degree of confidence, they would require the full gambit of input data that the defendant AI system had access to at the moment it made the inference or decision. This would require AI systems to record and store input data for all inferences, which may increase the total memory usage and cost for the system. Additionally, both AI system developers and users must agree to be transparent and willing to disclose the recorded data. Developers may not want to do so because they may have some proprietary interest in the input data collected by their products. Users may not want such data to be disclosed because it may contain potentially private information. This is a legitimate concern and limitation of the AI expert witness scheme, especially when considering the secrecy or inscrutability of black-box AI systems.[121]

121. *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 Fordham L. Rev. 1085, 1089-98 (2018) (Discussing the transparency concerns of

However, the benefits of documentation and transparency largely outweigh any detriments.[122] Documentation of input data collected shortly before a harm occurred would be highly beneficial to fact-finding; accurate records of the circumstances surrounding a harm may provide valuable testimony to the court. Disclosing all the input data collected would also increase public transparency; the public would have greater knowledge of exactly what types of input data are collected by the AI system and how much and often the data is collected. This data transparency is inherently valuable and perhaps necessary for any legal accountability of AI systems.[123] Such disclosure does not have to violate any privacy or proprietary data rights of the users and developers. Disclosure of input data can be done privately, with access only granted as needed for accountability.[124] Additionally, the memory cost of recording input data is not prohibitively burdensome and has been shown to be practicable in the automotive context.[125] Event data recorders are devices installed in motor vehicles which record and store data before, during, and after a crash including data on the vehicle's status and any driving decisions made by the driver.[126] NHTSA uses data from event data recorders to support crash investigations and has previously considered requiring the installation of event data recorders in all vehicles.[127] As the cost and size of memory storage steadily decreases every year, event data recorders seem increasingly plausible.[128] For autonomous vehicles, increasing the data stored in the data recorder to include input data is entirely practicable. Any small increase in cost due to more memory usage is likely justified by the benefits provided.

One final potential criticism of the AI expert witness proposal is that this liability scheme requires the existence of multiple AI systems that serve

---

black-box AI systems, that they may be secret, requiring specialized knowledge, inscrutable, and nonintuitive).

122. *Id*. at 1118-25 (Introducing transparency to black-box algorithms is valuable because it is an inherent good, it enables action, and exposes a basis for evaluation of the algorithm).

123. Margot E. Kaminski, *Understanding Transparency in Algorithmic Accountability*, *in* CAMBRIDGE HANDBOOK OF THE LAW OF ALGORITHMS 121, 121 (Woodrow Barfield ed., 2020) ("Transparency is necessary, if not sufficient, for building and governing accountable algorithms.").

124. Selbst & Barocas, *supra* note 121, at 1130-1135 (To preserve data rights, disclosure and examination necessary for transparency can be undertaken in private contexts such as in depositions or interrogatories. The GDPR – EU's privacy regulation – conditions such types of assessments on the implementation of "appropriate technical and organizational measures" to protect data subject rights.).

125. *See Event Data Recorder,* NAT'L HIGHWAY TRAFFIC SAFETY ADMIN. (last visited May 12, 2022), https://www.nhtsa.gov/research-data/event-data-recorder.

126. *Id*.

127. *Id*.

128. *See generally* Robert E. Fontana & Gary M. Decad, *Moore's Law Realities for Recording Systems and Memory Storage Components: HDD, Tape, NAND, and Optical*, 8 AIP ADVANCES 056506 (2018) (Discussing Moore's law which stipulates that transistor density, and hence, memory capacity, increases at a steady rate).

the same function within the same industry while being developed with different algorithms or different training data. For new and emerging AI markets, this is an entirely valid limitation on the AI expert witness system. If there is only one AI system used in a particular context and no competitors have been developed yet, there is no possibility for any AI expert witnesses to provide testimony for potential harms caused by the AI system. In fact, for new AI uses, what constitutes a harm may not even be sufficiently understood to enable regulatory entities to establish degree of confidence standards.

However, the AI market is fast-growing and competitive.[129] As new AI uses develop and spread the potential for harm increases, but this also opens the door to the development of a large variety of AI competitors which may serve as a sufficient pool of AI expert witnesses. The low overhead costs to the development of software may be prohibitive for any monopolistic AI market as the startup threshold is less of a barrier to entering the market. Thus, although the AI expert witness liability scheme may be inapplicable during the infancy of an AI use, it becomes applicable when the AI use becomes more widespread and the potential for harm increases accordingly.

CONCLUSION

In conclusion, the nature of AI decision-making is inherently different from human decision-making, rendering legal tools designed to assess human liability inappropriate when applied to AI systems. However, the need for new AI-centered legal tools presents the unique opportunity to create clear and strict quantitative standards centered around assessed degrees of confidence that AI systems must comply with to avoid liability. When paired with AI expert witnesses which may audit autonomous vehicles in a comprehensive and quantitatively appropriate manner, this new regime of AI legal tools presents an efficient and effective system for assessing AI liability without being subject to any pitfalls presented by the inadequate analogy between AI and human decision-making and knowledge.

129. *See Artificial Intelligence (AI) Market: 19.84% Y-O-Y Growth Rate in 2021*, TECHNAVIO (Mar. 21, 2022) ("The Artificial Intelligence (AI) Market Share is expected to increase by USD 76.44 billion from 2020 to 2025, with an accelerated CAGR of 21%.").