Algorithmic Harm and Design Defects: The Problem with Applying Design Defect Liability Standards to Machine Learning Products

Vincent E. Molinari[*]

[*] J.D. Candidate, 2024, Seton Hall University School of Law. My most heartfelt thanks to Professor Timothy Glynn, whose guidance and support was instrumental in the publication of this Comment. I would also like to express my sincerest gratitude to Allen Yang and Rami A. Bazoqa, without whose technical knowledge this Comment would not have been possible. Finally, thank you to my partner, Penelope M. Way, for serving as a continuing source of inspiration.

I. INTRODUCTION

From its inception, machine learning has revolutionized virtually every industry where it can be applied.[1] The myriad uses of machine learning and its derivations have led to its incorporation into many facets of daily life for both consumers and corporations, even where not explicitly obvious.[2] Recently, an ever-increasing number of consumer products have begun to incorporate this technology in an effort to improve operational efficiency while providing an improved experience for end-users.[3] Like all products, however, consumer products incorporating machine learning elements have the potential to cause consumers harm during use.[4]

Current standards and tests for assessing design defect liability in such cases, as they exist in their current form under the *Restatement (Third) of Torts* (hereinafter "*Restatement (Third)*"), do not adequately contemplate and account for the existence and functionality of such products. The risk-utility test assesses design defect liability by balancing the utility of a product design against alternative designs in

1 *See* Michael Evans, *The Machine Learning Revolution*, FORBES (Oct. 20, 2018, 11:15 AM), https://www.forbes.com/sites/allbusiness/2018/10/20/machine-learning-artificial-intelligence-could-transform-business/?sh=7a78c428c6c3 (discussing the application of machine learning technologies in various industries, emphasizing its applicability regardless of company size and form).

2 *See generally An On-Device Deep Neural Network for Face Detection*, APPLE: MACH. LEARNING (Nov. 2017), https://machinelearning.apple.com/research/face-detection (discussing the use of computer vision, a form of machine learning, in the iPhone's Face ID feature); *see also What is Machine Learning? A Definition*, EXPERT.AI (Mar. 14, 2022), https://www.expert.ai/blog/machine-learning-definition/ (providing examples of machine learning systems currently used by businesses, including chatbots used for customer support and healthcare systems designed to improve patient outcomes).

3 *See* Marita Zorotovich & Marty Donovan, *Current Use Cases for Machine Learning in Retail and Consumer Goods*, MICROSOFT: BLOG (Sep. 9, 2018), https://azure.microsoft.com/en-us/blog/current-use-cases-for-machine-learning-in-retail-and-consumer-goods/ (discussing various use cases for machine learning in consumer products and explaining that a need for operational efficiency and customer service drives utilization of the technology); TBRC Bus. Rsch., *The Increased Use of Machine Learning and Artificial Intelligence is Expected to Fuel the Digital Transformation Market*, GLOBENEWSWIRE (Sep. 14, 2022, 11:30 AM), https://www.globenewswire.com/news-release/2022/09/14/2516223/0/en/The-Increased-Use-Of-Machine-Learning-And-Artificial-Intelligence-Is-Expected-To-Fuel-The-Digital-Transformation-Market-As-Per-The-Business-Research-Company-s-Digital-Transformatio.html (discussing the recent rise in machine learning technology utilization.)

4 *See, e.g.*, *California Teenager Dies in Self-Driving Tesla Crash*, ENJURIS, https://www.enjuris.com/blog/news/tesla-autopilot-accident/ (discussing the death of a teenager caused by a malfunction in the machine learning-based aspect of a Tesla vehicle).

the context of foreseeable risks of harm stemming from the product's design.[5]

This Comment advances two arguments. First, a design defect standard incorporating the concept of foreseeability is flawed when applied to machine learning due to complications caused by covariate shift and concept drift, which have the potential to create harms unforeseeable to the designer from the foreseeable uses of the product. Second, this Comment argues that the reality of probability-based reasoning in machine learning renders the policy undergirding the current standard inapplicable due to difficulties in applying the reasonable alternative design standard to machine learning algorithms.

This Comment proceeds in four parts. First, given the relative complexity of machine learning technology, Part II will provide a brief non-technical explanation of the creation and functionality of machine learning, while also distinguishing its functionality and use from other forms of technologies. Second, Part III will examine the current legal awareness of machine learning and related legal issues through cases and academic materials. Drawing on this background, Part IV will identify the chief difficulties in applying current design defect standards to machine learning, focusing on the legal standard's clash with the practical reality of machine learning's functionality.

Finally, Part V will reflect on the prior discussion to provide a set of three conclusions for the identified legal issues, each of which strikes a different balance between the value one places on machine learning technology and adherence to existing design defect standards. The first conclusion, reflecting a low valuation of machine learning technology, posits that including a machine learning element in a product may represent a manifestly unreasonable design due to the identified issues with the technology, and analyzes the practical litigation effects of such an approach. The second, striking a balance between machine learning and existing design defect standards, advances an approach based on the insertion of an independent element of foreseeable risk into the design defect standard and a materiality requirement in reasonable alternative design proofs presented, both of which combined remedy the identifiable issues in the application of the law to machine learning products. The third and final conclusion, representing an exceedingly high valuation of machine learning, proposes that the lack of a machine learning element in some products may constitute a manifestly

5 *See* RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(b) (AM. L. INST. 1999).

unreasonable design due to the advantages of machine learning's use, thereby inversing the position of the first conclusion.

II. TECHNICAL BACKGROUND OF MACHINE LEARNING

Discussion of a highly technical topic, like machine learning in relation to specific areas of the law, requires an initial understanding of the technology itself. The goal of this section is twofold: first, to provide a concise definition of machine learning and a non-technical discussion of its developmental stages; second, to distinguish machine learning's use and functionality from other forms of computer technology.

A. Machine Learning Defined and the Process of Creation

As a threshold matter, it is important to note that the term "machine learning" is essentially a term of art used to reference various computational methods and technologies.[6] The broadest definition of the term possible, covering each of these computational methods and technologies, defines "machine learning" as a process that teaches computers through the "use of data and algorithms to imitate the way that humans learn."[7] This training allows a system "to automatically [spot] patterns in . . . data that can [then] be used to make predictions."[8] The key, however, is that machine learning seeks to automate away the need for a designer to explicitly program or instruct a system on how to make its predictions, instead allowing the system to learn for itself the most efficient and effective method by which to perform its function.[9]

This definition begs the question of how a designer may accomplish the lofty goal of creating a system that learns by itself. On this point, the

6 *See* Jason Brownlee, *14 Different Types of Learning in Machine Learning*, MACH. LEARNING MASTERY (Nov. 11, 2019), https://machinelearningmastery.com/types-of-learning-in-machine-learning/ (providing a list of the types of machine learning). A discussion of each individual type of machine learning is beyond the scope of this Comment but note that each type of machine learning has its own inherent advantages and disadvantages.

7 *What is Machine Learning*, IBM, https://www.ibm.com/cloud/learn/machine-learning (last visited Jan. 16, 2024).

8 Elizabeth Quirk, *Artificial Intelligence Umbrella Glossary: Machine Learning, AI, RPA & More*, SOLUTIONS REV. (Jan. 26, 2018), https://solutionsreview.com/business-process-management/artificial-intelligence-umbrella-glossary-machine-learning-ai-rpa/.

9 *See* EXPERT.AI, *supra* note 2 (explaining the focus of machine learning from an implementation perspective).

development of machine learning can be broken into seven basic stages, each of which bears strongly on the final product.[10]

The first step in developing any machine learning system entails the collection of data.[11] The designer must focus on finding a reliable data source that the completed system can utilize efficiently in its function.[12] Data collection may occur in several ways, such as the wholesale importation of data from an existing commercial database or the creation of a fresh dataset gathered and combined by a designer.[13] The designer must also ensure that the data is not irreparably flawed in some fashion, which may occur when data is outdated or incorrectly collected, among other reasons.[14]

Next, the designer must prepare the data.[15] This step focuses on randomizing the previously collected data to ensure that the ordering of the data does not affect the system's learning process.[16] The ordering of data is important to a machine learning system because it is important for a human student: the order in which one presents concepts affects the assumptions extrapolated from those concepts.[17] After

10 *See* Yufeng Gao, *The 7 Steps of Machine Learning*, MEDIUM: TOWARDS DATA SCI. (Aug. 31, 2017), https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e (discussing the steps of machine learning development and implementation). Note that these stages, or steps, may be more or less numerous depending on the context of the system, but the core elements of the following stages are always present.

11 *See* Mayank Banoula, *Machine Learning Steps: A Complete Guide*, SIMPLILEARN, https://www.simplilearn.com/tutorials/machine-learning-tutorial/machine-learning-steps (last visited Jan. 16, 2024).

12 *See id.* ("Make sure you use data from a reliable source as it will directly affect the outcome of your model.").

13 *See* Raul V. Rodriguez, *The 7 Key Steps to Build Your Machine Learning Model,* AI MYSTERIES (May 29, 2020), https://analyticsindiamag.com/the-7-key-steps-to-build-your-machine-learning-model/ ("You may have the information in an existing database or you must create it from scratch.").

14 *See* Banoula, *supra* note 11 ("The quality of data . . . will determine how accurate [the] model is. If [a designer] has incorrect or outdated data, [he] will have wrong outcomes or predictions which are not relevant.").

15 Gao, *supra* note 10.

16 *See* Banoula, *supra* note 11 (discussing the randomization of data and its importance to the learning process); Gao, *supra* note 10 (explaining that ordered data affects the machine learning process).

17 Gao, *supra* note 10 (discussing ordering of training data). This principle is embedded in the human mind and machine learning systems. For an example of this occurring in the human mind, see Eva Fourakis & Jeremy Cone, *Matters Order: The Role of Information Order on Implicit Impression Formation*, SOC. SCI. & PERS. SCI., Jan, 2020, at 56, 56–7 (discussing how the order an individual learns the personality traits of an unknown person affect the individual's perception of that person) [https://doi.org/10.1177/1948550619843930].

randomization, the designer must audit and prune the dataset to remove any errant redundancies or minor flaws that may be present.[18] Once audited, the collected data must be split into two distinct datasets: the training and testing data.[19]

Training data is the dataset used to train the system's predictive function.[20] It is helpful to think of it as a driver's education training course that teaches a student how to act behind the wheel. In contrast, testing data is the dataset used to evaluate the model's performance under real-world conditions.[21] Similarly, testing data is more akin to the driving test a student driver needs to pass before being allowed on the road. It is exceedingly important that the designer carefully analyzes the training and testing data, for if either dataset is overly broad or biased towards a specific variable, the system will not be able to develop an accurate predictive model.[22] This step, in its entirety, has critical implications for the rest of the creation process because a machine learning algorithm's predictive model is only as good as its data.[23] Quality training data is to a budding machine learning system as a quality casebook is to a law student: without a reliable basis of knowledge, there is only so far one can go.[24]

The third stage of development is where the designer chooses a machine learning model to implement.[25] Essentially, this step asks the designer to identify the precise problem they are attempting to solve

18 *See* Banoula, *supra* note 11 ("Cleaning the data to remove unwanted data, missing values, rows, and columns, duplicate values, data type conversion, etc.").

19 *See* Banoula, *supra* note 11 ("Splitting the cleaned data into two sets – a training set and a testing set."); *see also* Gao, *supra* note 10 ("Split the data into two parts," the training and testing data.).

20 *See* Banoula, *supra* note 11 ("The training set is the set your model learns from); *see also* Gao, *supra* note 10 (discussing the uses of training data).

21 *See* Banoula, *supra* note 11 ("A testing set is used to check the accuracy of your model after training.").

22 *See* Rodriguez, *supra* note 13 (emphasizing the importance and methodology of preparing data); Gao, *supra* note 10 (discussing the importance of the reliability of training data). This principle is similarly identifiable in the human mind as well. *See* Isabel Bilotta et al., *How Subtle Bias Infects the Law*, 15 ANN. REV. L. & SOC. SCI. 227, 230 (2019) (explaining the results of a study of the effects of racial bias in weapon identification) [https://doi.org/10.1146/annurev-lawsocsci-101518-042602].

23 *See* Rachel Wolff, *What is Training Data in Machine Learning*, MONKEYLEARN BLOG (Nov. 2, 2020), https://monkeylearn.com/blog/training-data/ (explaining that training data determines "just how smart [a] model can become.").

24 *See id.*

25 *See generally* Rodriguez, *supra* note 13 (providing a non-exhaustive list of machine learning models for potential selection and explaining the problem each seeks to solve).

and choose a machine learning model best suited to solve it.[26] For example, if a designer attempts to predict an outcome based on a class or category of related inputs, it is best to use an algorithmic classification model.[27] In contrast, a regression model is better suited to predict an outcome based on a set of independent variables,[28] such as attempting to determine the relationship between "employee satisfaction and product sales."[29] Designers can choose from various models, each tailored to solve a specific type of problem.[30]

The next stage entails the bulk of the work: training the designer's selected model.[31] The previously prepared training dataset is passed to the system to allow it to attempt its predictive function.[32] The system then runs the selected predictive model on the training data repeatedly, iteratively improving its predictions based on its interpretation of the dataset and possible micro-adjustments the designer makes to the system.[33] In the first iterations of training, the system will essentially make random predictions with no real methodology or practice, producing highly inaccurate and random results.[34] However, over many training cycles, the system will iteratively improve its predictive model to the point where it can produce accurate predictions from the data provided.[35]

---

26 *See generally* Rodriguez, *supra* note 13.

27 *See* Natassha Selvaraj, *8 Machine Learning Models Explained in 20 Minutes,* DATACAMP BLOG (Sep. 2020), https://www.datacamp.com/blog/machine-learning-models-explained (explaining that a classification algorithm is best suited to solve the problem of predicting heart disease on a number of risk factors).

28 *See id.* (providing that a regression model is best to predict the rent of a house based on a number of independent factors).

29 Catherine Cote, *What is Regression Analysis in Business Analytics?*, HARVARD BUS. SCH. ONLINE (Dec. 14, 2021), https://online.hbs.edu/blog/post/what-is-regression-analysis.

30 Two examples of machine learning models are provided to illustrate the selection process. A discussion of each individual model and its related benefits is beyond the scope of this Comment.

31 *See* Gao, *supra* note 10.

32 *See* Banoula, *supra* note 11 ("In training, you pass data to [the] machine learning model to find patterns and make predictions.").

33 *See* Rodriguez, *supra* note 13 (discussing incremental improvement of the predictive model based on repeated testing cycles); *see also* Gao, *supra* note 10 (explaining in detail the training process of a hypothetical machine learning system).

34 *See* Gao, *supra* note 10 ("When [the designer] first starts the training, it's like [the system] drew a random line through the data.").

35 *See* Gao, *supra* note 10 ("[A]s each step of the training progresses, the line moves, step by step, closer to [an accurate prediction]."); Banoula, *supra* note 11 ("Over time, with training, the model gets better at predicting."). A discussion on the precise method by which machine learning learns is a highly technical topic beyond the scope of this Comment. For a technical discussion of the process, see M. I. Jordan & T. M. Mitchell,

After training is complete, the designer must evaluate the system.[36] The previously reserved testing dataset serves as the benchmark for this evaluation.[37] The idea is to evaluate how the system reacts to data it has not previously observed, representing its expected functioning in real-world conditions after its completion and implementation.[38] This step is necessary to ensure the system's functionality before its implementation in real-world conditions because a high degree of predictive accuracy on training data does not generally indicate an accurate predictive model.[39] As an illustration, a student who rigorously and exclusively studies for his Constitutional Law exam throughout the semester may achieve a high grade on that exam, but the same cannot be said for his Evidence exam.

Once evaluation shows that the system performs well on unseen data, the designer may find it appropriate to tune the parameters of the model to achieve a higher degree of accuracy.[40] To accomplish this, a designer can adjust specific controlled variables within the system and reiterate the training and testing cycles to determine if a more accurate prediction is possible.[41] At some undetermined value of each controlled variable, the accuracy of the predictive model will be at its peak. Parameter tuning helps the designer determine the values that will allow the system to reach peak accuracy.[42]

At the final step, the model is ready for implementation.[43] At this point, the machine learning system may be provided unseen, real-world data and is expected to make accurate predictions using the predictive

Machine Learning: Trends, Perspectives, and Prospects, 349 SCI. 255, 257–60 (2015) [https://doi.org/10.1126/science.aaa8415].

36 *See* Rodriguez, *supra* note 13.

37 *See* Gao, *supra* note 10 ("This is where the [the testing data] that we set aside earlier comes into play.").

38 *See* Banoula, *supra* note 11 ("This is done by testing the performance of the model on previously unseen data. The unseen data used is the testing set that you split our data into earlier."); Gao, *supra* note 10 ("This is meant to be representative of how the model might perform in the real world.").

39 *See* Banoula, *supra* note 11 ("If testing was done on the same data which is used for training, [the designer] will not get and accurate measure, as the model is already used to the data, and finds the same patterns in it, as it previously did. This will [show] disproportionately high accuracy.").

40 *See* Gao, *supra* note 10.

41 *See* Gao, *supra* note 10 (explaining that assumed control variables may be tuned to achieve higher accuracy); Banoula, *supra* note 11 ("Parameters are the variables in the model that the programmer generally decides.").

42 *See* Banoula, *supra* note 11 (discussing the broader goals of parameter tuning).

43 Rodriguez, *supra* note 13.

model developed through the prior steps.[44] Training, testing, and parameter tuning lead to this point: where "the value of machine learning is realized."[45] At this stage, the designer can implement the system for its purpose, and the system can be trusted to make accurate predictions with its developed model.[46]

B. Machine Learning Distinguished

While understanding a concise definition of machine learning and the process by which it is created is helpful, it is far more important to distinguish machine learning from other, more traditional forms of technology to understand its use best.

The key difference to appreciate between machine learning and other forms of traditional computational technology is the difference in the problem that each attempts to solve. Most computational systems aim to solve a predefined problem, such as the translation of a document,[47] providing a defined output based on a set of rules programmed into the system.[48] In this way, the system is rules-based and expert-driven, meaning that it mimics the knowledge of an expert on the topic and is thus capable of producing an output similar to that of an expert with the same amount of information.[49] An exceedingly simple example of a rules-based system, as described, is the common four-function calculator;[50] the calculator takes the numbers typed in, the data, and transforms them into an output based on the predefined functions encoded into the system.[51] Therefore, rules-based systems

44 *See* Banoula, *supra* note 11 ("... [Y]ou can use [the] model on unseen data to make predictions accurately.").

45 Gao, *supra* note 10.

46 Gao, *supra* note 10.

47 *See* ROUTLEDGE ENCYCLOPEDIA OF TRANSLATION TECHNOLOGY 454 (Sin-Wai Chan ed., 2d ed. 2015) ("Traditionally, [translation] systems use either a rules-based or corpus-based approach to translate a document.").

48 *See* Vasudevan Swaminathan, *The Conundrum of Using Rule-Based vs. Machine Learning Systems*, ZUCI SYSTEMS BLOG, https://www.zucisystems.com/blog/the-conundrum-of-using-rule-based-vs-machine-learning-systems/ ("Rule-based systems are computer programs that use if-then rules to make decisions and perform tasks.").

49 *See id.* ("Human experts build rule-based systems with in-depth domain knowledge to guarantee the best possible outputs. Hence, they are expert-driven systems.").

50 *See, e.g.*, DESMOS, https://www.desmos.com/fourfunction (last visited Feb. 4, 2023).

51 *See, e.g.*, Natalie Wolchover, *How Do Calculators Calculate?*, LIVESCIENCE (May 10, 2011), https://www.livescience.com/14087-calculators-calculate.html (explaining the computer logic of how a calculator produces an output).

only mimic the knowledge of an expert to aid in the decision-making process, rather than replace the expert entirely.[52]

In contrast, a machine learning system is used to analyze data patterns that facilitate the rules' development.[53] Therefore, the distinguishing feature of machine learning systems is that they do not require manual programming to an end, as the entire purpose of the system is to avoid the need for a human programmer to define the rules by which the system operates.[54] It is clear, then, that the use cases of machine learning and rules-based systems are entirely dissimilar because a rules-based system cannot analyze a problem in the same way expected of a machine learning system.[55] In this way, the difference between the two forms of technology can be understood as a distinction between a system designed to play a television show, and a system designed to find the *best* television show.[56] The former simply performs its function based on the inputs of the user, while the latter takes input data from the operator and transforms it into a set of rules to identify the quantifiably best result.[57]

---

52 For example, it cannot be realistically suggested that all mathematicians will simply be replaced with calculators. *See* Erez Yereslove, *Calculators Didn't Replace Mathematicians, and AI Won't Replace Humans*, WORLD ECON. F. (Jan. 29, 2019), https://www.weforum.org/agenda/2019/01/calculators-didnt-replace-mathematicians-ai-automation-work/ (suggesting that calculators only improved the mathematician's function, rather than replacing them altogether).

53 *See* Swaminathan, *supra* note 48 ("The important point to note here is that no one needs to tell the information . . . to the Machine Learning-based system. The software can make this logical deduction on its own by simply analyzing the data and looking for correlations.").

54 *See* Swaminathan, *supra* note 48 ("The biggest difference between rule-based systems and [machine learning] systems is that humans manually program rule-based systems, whereas machines automatically train self-learning systems. In other words, self-learning systems learn from experience rather than being explicitly told what to do by humans.").

55 *See* Bernard Marr, *The Top 10 AI and Machine Learning Use Cases Everyone Should Know About*, FORBES (Sep. 30, 2016), https://www.forbes.com/sites/bernardmarr/2016/09/30/what-are-the-top-10-use-cases-for-machine-learning-and-ai/?sh=46a145a094c9 (discussing the distinguishing aspect of machine learning and providing examples of its use in everyday life).

56 *See, e.g.*, Libby Plummer, *This is How Netflix's Top-Secret Recommendation System Works*, WIRED U.K. (Aug. 22, 2017, 7:00 AM), https://www.wired.co.uk/article/how-do-netflixs-algorithms-work-machine-learning-helps-to-predict-what-viewers-will-like (explaining the functionality of machine-learning recommendation engines used by Netflix).

57 The problem the rules-based system in this example seeks to solve is "How can I watch a television show?" which is answered by the functionality of the system allowing a user to press a button to play the program. The sole rule the system operates by is the television show playing once the play button is pressed. In contrast, the machine learning system is expected to develop a set of recommendations to the user based on

III. JUDICIAL AWARENESS AND SCHOLARLY COMMENTARY ON MACHINE LEARNING'S INTERSECTION WITH PRODUCTS LIABILITY

This section briefly describes the range of academic literature on the topic and a few cases that highlight the practical legal knowledge on the topic, or possible lack thereof, to understand the legal landscape related to machine learning.

A. Scholarly Commentary on Machine Learning

Much has been written on design defects and the relative merits of using one test or another to ascertain the liability of a designer or seller.[58] Unfortunately, the same cannot be said for the intersection of design defect law and machine learning, a topic that has thus far produced little scholarship.

When considering the published material in this space, the article *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision Makers* stands above the rest in its influence on the development of the field.[59] Focusing primarily on whether machine learning can be considered a product for product liability law, the article briefly discusses the difficulty of applying the concept of defective design to harm caused by the probability-based decision-making process of machine learning.[60] Although identifying some of the pertinent issues, this section of the article does not come to a definitive conclusion on the issues presented, leaving an open question as to whether changes in existing doctrine are necessary.[61] Thus, the article does not address the core question posed by this Comment, namely, how

collected data to answer the question "What should I watch?" For a discussion on the functionality of machine learning in recommendation systems, see Rohit Dwivedi, *What Are Recommendation Systems in Machine Learning?*, ANALYTIC STEPS (Apr. 16, 2020), https://www.analyticssteps.com/blogs/what-are-recommendation-systems-machine-learning (explaining how recommendation systems work and providing examples of their use).

58 *See* discussion *infra* Section IV.A and note 96.

59 Karni A. Chagal-Feferkorn, *Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision Makers*, 30 STAN. L. POL'Y REV. 61 (2019); *see also* Lauren Rhue & Anne L. Washington, *AI's Wide Open: Premature Artificial Intelligence and Public Policy*, 26 B.U. J. SCI. & TECH. L. 353, 369 (stating that substandard AI may "fall under . . . defective products legal theories."); Richard E. Kaye, *Distinguishing Between Products and Services, in* AMERICAN LAW OF PRODS. LIAB. 3D § 16:67, (2024) (distinguishing between products and services in the context of strict liability).

60 *See* Chagal-Feferkorn, *supra* note 59, at 84–87 (discussing probability-based harms in the context of medical algorithms).

61 *See id.* at 85–86 ("Does this mean that [machine learning] should never be governed by products liability and that our analysis should have nothing to do with said legal framework? Not necessarily.").

to reconcile current design defect standards with the core functionality of machine learning.

Other commentators have come similarly close to identifying the core issues of the intersection between design defects and machine learning but stop short of addressing the root causes and cures of these problems, opting to leave the complex topic as an open question.[62] Moreover, many of these pieces discuss the concepts of machine learning and product liability in only specific contexts, such as in the medical or healthcare fields.[63]

In contrast, some commentators focus on the metaphysical question of reasonableness in machine learning's decision-making ability.[64] Others analyze implicit bias in machine learning and its potential to cause harm when used by the government, attempting to determine the precise means by which liability should be imposed in such a situation.[65] Some anchor discussion of machine learning and its implications of the field as a whole, with divergences and references to substrata of negligence doctrine, to illustrate the inherent tension between machine learning and liability premised on human negligence.[66]

Thus, the existing jurisprudence on machine learning reads like a Jackson Pollock painting.[67] From afar, it seems like a jumbled mess: a grayscale spattering of un-connectedness. However, the common thread that can be appreciated from afar is that current scholarship does not answer the questions posed by this Comment, and that, by

62 *See, e.g.*, Vivian D. Wesson, *Who (Or What) is Liable for AI Risks?*, 92 N.Y. BAR J. 18, 20 (2020) (identifying foreseeable uses of a 3D printer as a problem for application of design defect theory).

63 *See* Sarah Kamensky, *Artificial Intelligence and Technology in Health Care: Overview and Possible Legal Implications*, 21 DEPAUL J. HEALTH CARE L. 1, 1–2 (discussing machine learning's use in the health care setting); *see also* Samuel D. Hodge, *The Medical and Legal Implications of Artificial Intelligence in Health Care—An Area of Unsettled Law*, RICH. J.L. & TECH. 405, 442 (discussing the drawbacks and complexities of "applying products liability law to AI in medical setting.").

64 *See, e.g.*, Karni Chagal-Feferkorn, *The Reasonable Algorithm*, 2018 U. ILL. J.L. & POL'Y 111, 121 (2018) (explaining a proposed reasonable algorithm standard and comparing its relative merits to other modes of liability).

65 *See, e.g.*, Christine Kumar, *The Automated Tipster: How Implicit Bias Turns Suspicion Algorithms into BBQ Beckys*, 72 FED. COMMC'N L. J. 97, 115–21 (2020) (providing two alternative modes of liability for harms caused by machine learning in the law enforcement context).

66 *See* Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. Rev. 1315, 1321–22 (2020) (asking whether "negligence law can successfully adapt to AI" and specifically discussing decision-assistance machine learning).

67 *See, e.g.*, *Number 5, 1948 by Jackson Pollock*, JACKSON POLLOCK, https://www.jackson-pollock.org/number-5.jsp.

extension, design defect liability standards are not equipped to address the unique issues posed by machine learning.

B. Judicial Awareness of Machine Learning

First and foremost, it must be noted that machine learning is a relatively new concept in case law.[68] Thus, awareness of the issues posed by machine learning is questionable at best, mostly due to the lack of established doctrine surrounding the technology.[69] However, there are indications of a shift in awareness of the technology, especially in recent years.

For example, in *Zaletel v. Prisma Labs, Inc.*,[70] a trademark infringement case, a court saw the use of machine learning as a highly differentiating factor between otherwise similar products.[71] The court noted the "very real differences in functionality" to conclude that "the two products are directed to different consumers."[72] Disregarding the admittedly narrow holding, this case can be viewed as judicial

68 The first opinion mentioning the term dates to 2002, a mere twenty-one years ago, but its impact on the case was negligible at best. American Library Ass'n, Inc. v. U.S., 201 F. Supp.2d 401, 433 (E.D. Pa. 2002) ("These algorithms sometimes make reference to the position of a word within text . . . [and] the weights are usually determined by machine learning methods."). Similarly, the term artificial intelligence was first mentioned in a 1988 opinion but was not discussed in any capacity until 2000. *In re* Estate of McCool, 553 A.2d 761, 763 (N.H. 1988) ("a corporation specializing in developing and marketing computer equipment and artificial intelligence software."); Qualitative Reasoning Systems, Inc. v. Computer Sciences Corp., No. 98CV554, 2000 WL 852127, at *1 (D. Conn. 2000) ("Traditional artificial intelligence programs employ 'fault trees' or 'rules-based logic' to diagnose failures.") (citation omitted). An attentive eye will notice the seemingly incorrect definition of artificial intelligence provided, attesting to the rapid and continuing development in this field of technology.

69 For example, copyright law is unclear on the question of whether the use of copyrighted materials in the training of a machine learning algorithm represents a copyright violation. *See* Cassandra Coyer, *Lawyers Expect More Litigation, and Clarity, Around Machine Learning's Copyright Issues*, LAW.COM (Aug. 19, 2022, 10:00 AM), https://www.law.com/legaltechnews/2022/08/19/lawyers-expect-more-litigation-and-clarity-around-machine-learnings-copyright-issues/ (speculating on the question of whether copyright is infringed when using copyrighted materials in machine learning datasets). Additionally, the Supreme Court recently heard a case on whether social media corporations should enjoy publisher immunity when using machine learning algorithms to direct users to content on their platforms. However, the Court did not reach the issue because it found that the plaintiffs did not state a claim. Gonzalez v. Google LLC, 143 S. Ct. 1191, 1191 (2023).

70 Zaletel v. Prisma Labs, Inc., No. 16-1307, 2017 U.S. Dist. WL 877302 (D. Del. Mar. 6, 2017).

71 *See id.* at *6 (explaining that "while plaintiff broadly describes both apps as 'photo filtering apps, the record demonstrates that defendant's app analyzes photos using artificial intelligence technology.").

72 *Id.*

recognition of machine learning's divergence from other types of computer technologies. If the court had disregarded the difference in functionality, there would have been a stronger case for an overlap between the products serving as a subject of the litigation.[73]

In a similar vein, the court in *Aerotek, Inc. v. Boyd* denied a motion for rehearing en banc,[74] which concerned the enforceability of an electronically signed arbitration agreement.[75] The dissent in that opinion expressly acknowledged that courts may one day have to determine whether machine learning algorithms have altered previously signed agreements to adjudicate a case properly.[76] While there was little discussion on this point,[77] the judicial acknowledgment of the capabilities and functionality of machine learning is a step closer to mainstream legal awareness on the topic.

In the tort law context, however, there has been an increase in discussion on machine learning resulting from the inclusion of autonomous driving features in automobiles.[78] The first known case alleging harm from the use of a machine learning autonomous technology was filed in 2018, in which a motorcyclist claimed that he sustained injuries after a General Motors autonomous vehicle merged into his lane and knocked him to the ground.[79] Interestingly, the plaintiff did not claim that the autonomous vehicle's operator contributed to the accident, naming only General Motors as a defendant.[80] The complaint's sole claim was negligence, alleging that General Motors "owed Plaintiff a duty of care in having its [autonomous vehicle] operate in a manner in

73 *See id.* (contrasting the differences between the apps and finding the use of machine learning as the main difference between them).

74 Aerotek, Inc. v. Boyd, 598 S.W.3d 373 (Tex. App. 2020).

75 *Id.* at 375.

76 *Id.* at 379 n.9 (Schenck, J., dissenting) (acknowledging that there was no evidence that the software at issue "became self-aware and rewrote the agreements," but noting that "machine learning and artificial intelligence may one day force us to confront these issues.").

77 *Id.* (Schenck, J., dissenting) ("I conclude the evidence was of no legal relevance, [and] I find further debate unnecessary.").

78 *See, e.g.*, Kayla Matthews, *Legal Implications of Driverless Cars*, L. TECH. TODAY (Oct. 3, 2018), https://www.lawtechnologytoday.org/2018/10/the-legal-implications-of-driverless-cars/ (questioning the legal implications of autonomous vehicles in tort litigation and suggesting that lawyers must adapt to the challenges). This is not to suggest that autonomous vehicles are the only means by which a machine learning algorithm can cause harm, but simply that it is likely the most frequent in today's technological landscape.

79 Complaint & Demand for Jury Trial at 2–3, Nilsson v. General Motors LLC, No. 18cv471 (N.D. Cal. Jan. 22, 2018).

80 *Id.* at 2.

which it obeys [] traffic laws and regulations."[81] Thus, rather than couching his claim in any product liability-related cause of action, the plaintiff attempted to litigate the claim similarly to a typical auto accident negligence claim.[82] General Motors admitted "that the [autonomous vehicle] was required to use reasonable care in driving,"[83] but the issue was never litigated due to the case settling four months after it was filed.[84]

Next, in March 2018, the use of autonomous driving technology in an Uber automobile led to the death of a pedestrian.[85] The inattentive backup operator of the automobile, tasked with monitoring the autonomous system, failed to stop the vehicle before it hit a pedestrian walking with a bicycle across the street.[86] The backup operator was criminally charged with negligent homicide for her role in the accident,[87] with the county attorney handling the case finding that there

81 *Id.* at 4.

82 While negligence may not be the only claim alleged in a typical automobile accident case, negligence usually forms the basis of the complaint. *See, e.g.*, Luciano v. Islam, 171 N.Y.S.3d 749, 754–55 (Sup. Ct. 2022) (discussing the elements of negligence in an automobile accident case in reference to the state's summary judgment standard).

83 Answer at 4, Nilsson v. General Motors LLC, No. 18cv471 (N.D. Cal. Jan. 22, 2018).

84 *See* RJ Vogt, *GM Settles First-Known Suit Over Self-Driving Car Crash*, LAW360 (June 1, 2018, 10:56 PM), https://www.law360.com/articles/1049776/gm-settles-first-known-suit-over-self-driving-car-crash ("According to court filings, counsel for both sides met and agreed in April to enter private mediation . . . [o]n May 30, they filed joint notice of settlement.").

85 *See* Andrew J. Hawkins, *Serious Safety Lapses Led to Uber's Fatal Self-Driving Crash, New Documents Suggest*, VERGE (Nov. 6, 2019, 11:45 AM), https://www.theverge.com/2019/11/6/20951385/uber-self-driving-crash-death-reason-ntsb-dcouments (discussing the details of the fatal accident and a related report from the National Traffic Safety Board).

86 *See* Phil McCausland, *Self-Driving Uber Car That Hit and Killed Woman Did Not Recognize That Pedestrians Jaywalk*, NBC NEWS (Nov. 9, 2019, 3:28 PM), https://www.nbcnews.com/tech/tech-news/self-driving-uber-car-hit-killed-woman-did-not-recognize-n1079281 ("... the car couldn't recognize [the victim] as a pedestrian or a person . . . [and the safety driver] was streaming the television show 'The Voice.'"). For the safety driver's differing account of the situation, see Lauren Smiley, '*I'm the Operator': The Aftermath of a Self-Driving Tragedy*, WIRED (Mar. 8, 2022, 6:00 AM), https://www.wired.com/story/uber-self-driving-car-fatal-crash/.

87 *See* Katyanna Quach, *Driver in Uber's Self-Driving Car Death Goes on Trial, Says She Feels Betrayed*, THE REGISTER (Mar. 14, 2022, 1:19 PM), https://www.theregister.com/2022/03/14/in_brief_ai/ (discussing the backup driver's perspective on the negligent homicide charge). At the time of this Comment's publication, the backup driver has pleaded guilty to endangerment and sentenced to three years of supervised probation. Corina Vanek, *Arizona Driver in Fatal Autonomous Uber Crash in 2018 Pleads Guilty, Sentenced to Probation*, AZCENTRAL, https://www.azcentral.com/story/news/local/tempe/2023/07/28/rafaela-vasquez-pleads-guilty-in-in-fatal-uber-self-driving-crash-killed-pedestrian-elaine-herzberg/70488361007/ (July 28, 2023, 8:29 A.M.)

was no basis for the criminal liability of Uber itself.[88] Uber, however, quickly settled a related civil suit brought by the deceased pedestrian's family out of court, thereby declining to litigate the issue.[89]

Since these two initial cases, multiple complaints have been filed against autonomous vehicle manufacturers, each of which usually alleges a design defect cause of action.[90] Of the cases still being litigated, the factual background of *Hinze v. Tesla*[91] may represent the best opportunity for doctrinal developments in the field of machine learning torts.[92] The complaint, which alleges a product liability cause of action among others,[93] includes a detail not mentioned in many other complaints: the plaintiff was actively monitoring the vehicle while it exercised its autonomous function.[94] This detail is significant because it may force litigation to focus more on the products liability theory presented by the plaintiff, as driver fault may be difficult to prove on the manufacturer's part. Therefore, the case has the potential to focus more on the allegedly defective design of the machine learning aspects of the automobile than in prior cases,[95] which in turn may coalesce into a

88 Letter from Sheila Polk, Yavapai County Attorney, to The Honorable Bill Montgomery (Mar. 4, 2019), https://s3.documentcloud.org/documents/5759641/UberCrashYavapaiRuling03052019.pdf.

89 *See* Scott Neuman, *Uber Reaches Settlement with Family of Arizona Woman Killed By Driverless Car*, NAT'L PUB. RADIO (Mar. 29, 2018, 3:23 AM), https://www.npr.org/sections/thetwo-way/2018/03/29/597850303/uber-reaches-settlement-with-family-of-arizona-woman-killed-by-driverless-car ("Uber Technologies has reached a settlement with the family of the woman killed earlier this month.").

90 While an in depth analysis of each individual case is beyond the scope of this Comment. *See, e.g.*, Complaint & Jury Demand, Banner v. Tesla Inc., No. 2019CA009962 (Fla. Cir. Ct. Aug. 1, 2019) (alleging a products liability cause of action against an autonomous vehicle manufacturer); *see also* Isobel A. Hamilton, *'We Cannot Have Technology and Sales Take Over Safety': Tesla Is Being Sued Again for a Deadly Autopilot Crash*, BUS. INSIDER (Aug. 2, 2019, 7:56 AM), https://www.businessinsider.com/tesla-sued-family-jeremy-beren-banner-autopilot-crash-2019-8 (discussing the factual background of Banner's death and summarizing prior lawsuits on the same issue).

91 Hinze v. Tesla, Inc., No. 22cv2944, (N.D. Cal. Apr. 4, 2022).

92 Complaint & Demand for Jury Trial at 1–7, Hinze v. Tesla, Inc., No. 22cv2944, (N.D. Cal. Apr. 4, 2022).

93 *Id.* at 13.

94 *See id.* at 2 ("Plaintiff was actively and consciously maintaining active supervision of the vehicle."); *see also* Joseph Geha, *Lawsuit: Tesla Autopilot Accelerated on Its Own, Causing Crash*, GOV'T TECHNOLOGY (May 20, 2022), https://www.govtech.com/fs/lawsuit-tesla-autopilot-accelerated-on-its-own-causing-crash ("The suit alleges that, unlike some other Tesla crashes involving the autopilot feature, [plaintiff] was actively and consciously maintaining active supervision of the vehicle.").

95 *See, e.g.*, Jonathan Stempel, *Jury Finds Tesla One Percent Negligent in Fatal Model S Crash*, REUTERS (July 19, 2022, 7:18 PM), https://www.reuters.com/business/autos-

usable framework for later cases where machine learning allegedly causes harm.

In conclusion, there is little to no case law speaking to the issues addressed in this Comment. Future litigation may lead to the development of doctrine addressing the intersection of machine learning and products liability, but the law has yet to develop to that end.

IV. APPLICATION OF DESIGN DEFECT STANDARDS TO MACHINE LEARNING PRODUCTS

Given the unresolved issues arising from the intersection of machine learning and product liability law, this section will proceed in three parts. Part A will give a brief justification of this Comment's choice to exclusively analyze the identified issues within the framework provided by the *Restatement (Third)*. Part B will provide a primer on the applicable law and its policy-based goals. Finally, Part C will discuss specific issues in applying the doctrine to products incorporating a machine learning element.

A. Justification for the Restatement (Third) Approach

This Comment analyzes the subsequent identifiable doctrinal issues in the context of the *Restatement (Third)* rather than through the approaches of any single state or prior formulations of design defect doctrine. Although controversial in both the academic and practical legal contexts at its inception,[96] the risk-utility test and the reasonable

transportation/jury-finds-tesla-just-1-liable-owes-105-mln-over-fatal-crash-2022-07-19/ (discussing an action brought against Tesla which turned on issues of comparative fault among the parties).

96 A variety of changes from the prior Restatement led to controversies, none more so than the changes to RESTATEMENT (SECOND) OF TORTS § 402A (AM. L. INST. 1965). The abandonment of prior doctrine in favor of the risk-utility test was a highly contested issue during the Restatement project itself, and the debates continued into the courtroom. *See* Victor E. Schwartz, *The Restatement (Third) of Torts: Products Liability: A Guide to Its Highlights*, 34 TORT & INS. L. J. 85, 88 (1998) (stating that "there was no issue that brought about more debate in the entire *Restatement* project than" the adoption of the risk-utility test and related alternative design proof requirements); John F. Vargo, *The Emperor's New Clothes: The American Law Institute Adorns a "New Cloth" for Section 402A Products Liability Design Defects—A Survey of the States Reveals a Different Weave*, 26 U. MEM. L. REV. 493, 502–03 (1996) (broadly criticizing the methodology of the creation of the risk-utility test and suggesting that the prevailing consensus behind the rule may be incorrect); Potter v. Chicago Pneumatic Tool Co., 694 A.2d 1319, 1322 (Conn. 1997) (declining to adopt the *Restatement (Third)* approach to design defects because "the feasible alternative design requirement imposes an undue burden on plaintiffs.").

alternative design requirement proposed by the *Restatement (Third)* has become the dominant approach among state courts wrestling with design defect litigation.[97]

In the few states that have outright rejected the risk-utility test, the main objection seems to be the insertion of a negligence-like standard into a doctrine governed by strict liability,[98] referring to the insertion of language focusing on the "foreseeable risks of harm posed by the product."[99] South Dakota, for example, did not outright reject the test; it simply declined to clarify whether the state followed the risk-utility test.[100] However, in states that have not adopted the *Restatement (Third)* approach, qualifications to the prior approach have brought the doctrine closer to the risk-utility test in some specific factual situations.[101] Additionally, South Dakota's unclear approach is likely to be clarified in favor of the *Restatement (Third)* approach once a case presents a favorable opportunity to do so.[102] Thus, given the *Restatement (Third)'s* dominance and broader influence, this Comment focuses on the risk-utility approach to design defects adopted by the *Restatement (Third)* .

---

97 The *Restatement (Third)* itself notes only six states that continue to utilize *the Restatement (Second) of Torts* consumer-expectations test as an independent standard, and one state that utilizes a mixed approach depending on the product at issue. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(b) reporter's note cmt. d (AM. L. INST. 1999).

98 *See* Ford Motor Co. v. Trejo, 402 P.3d 649, 656 (Nev. 2017) (the risk-utility test "inserts a negligence standard into an area of law where this court has intentionally departed from traditional negligence analysis."); Aubin v. Union Carbide Corp., 177 So.3d 489, 510 (Fla. 2015) (the risk-utility test "is inconsistent with the rationale behind the adoption of strict products liability.").

99 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(b) (AM. L. INST. 1999); *see also* Ford Motor Co. v. Trejo, 402 P.3d at 656 ("Rather than focus on the product itself, the risk-utility test subverts this analysis, focusing on the 'foreseeable risks of harm' apparent to the manufacturer when adopting the design.").

100 Karst v. Shur-Co., 878 N.W.2d 604, 610 n.4 (S.D. 2016) ("The dissent argues that we should adopt the risk-utility balancing test . . . [but] without the benefit of briefing and argument, we must wait for an appropriate case to consider such significant changes in our products-liability jurisprudence.").

101 *See* Cavanaugh v. Stryker Corp., 308 So.3d 149, 155–156 (Fla. Dist. Ct. App. 2020) (distinguishing *Aubin* on the basis that "some products may be too complex for a logical application of the consumer expectations test" and explaining that "the relevant expectations [for a complex medical device] are those of the medical professional, not the ordinary consumer."); *see also* Traci T, McKee, *Florida Appellate Court Authorizes the Use of the Risk-Utility Test in Complex Medical Device Cases*, FAEGRE DRINKER ON PRODS. (October 16, 2020), https://www.faegredrinkeronproducts.com/2020/10/florida-appellate-court-authorizes-the-use-of-the-risk-utility-test-in-complex-medical-device-cases/ (discussing the rejection of the consumer expectation test in *Cavanaugh*).

102 *See* Karst, 878 N.W.2d at 622 (Kern, J., dissenting) ("It is clear from the arguments of counsel that our sole reliance on the outdated principles contained in the Restatement (Second) of Torts § 402A is no longer workable.").

B. The Third Restatement's Design Defect Standard

As alluded to in prior sections, the *Restatement (Third)* uses what has been coined as "the risk-utility test" to determine when a product is defective in design.[103] Specifically, a product "is defective in design 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."[104] The risk-utility test functions as a balancing test, considering "a broad range of factors" such as the "magnitude and probability of foreseeable risks of harm, the instructions and warnings accompanying the product, and the nature and strength of consumer expectations regarding the product."[105] To prove a design defect, a plaintiff must present evidence of an available reasonable alternative design and then demonstrate a positive balancing of the risk-utility factors based on that alternative design.[106]

The underpinning of the test rests on the concept of foreseeability. There are two aspects of foreseeability at play: foreseeability of use and foreseeability of harm.[107] Foreseeability of use may be thought of as the likely results of the regular use of a product or a reasonably anticipated

---

103 In contrast to manufacturing defects, design defects require "an independent assessment of advantages and disadvantages, to which some attach the label 'risk utility balancing.'" RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. a (AM. L. INST. 1999).

104 *Id.* at § 2(b).

105 *Id.* at § 2 cmt. f. The list of factors provided in the comment to the Restatement (Third) is not meant to be exhaustive. *See, e.g.*, Banks v. ICI Americas, Inc., 450 S.E.2d 671, 675 n.6 (Ga. 1994) (remarking that "[n]o finite set of factors can be considered comprehensive or applicable under every factual circumstance," and providing a "non-exhaustive list of general factors."). The Wade factors have become an authoritative list of factors, even though they are not strictly applied by many courts. John W. Wade, *On the Nature of Strict Tort Liability for Products*, 44 MISS. L.J. 825, 837–38 (1973) (providing a list of seven factors to be used in risk-utility analysis). For a criticism of the Wade factors, see W. Kip Viscusi, *Wading Through the Muddle of Risk-Utility Analysis*, 39 AM. U. L. REV. 573, 580–81 (1990).

106 *See, e.g.*, Genie Industries, Inc. v. Matak, 462 S.W.3d 1, 9–10 (Tex. 2015) (suggesting that evidence of a reasonable alternative design is a prerequisite to the submission of risk-utility factors to the jury); *accord* RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. D (AM. L. INST. 1999) ("Assessment of a product design in most instances requires a comparison between an alternative design and the product design that caused the injury, undertaken from the viewpoint of a reasonable person.").

107 *Compare* RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. M (AM. L. INST. 1999) ("[The risk-utility test will] impose liability only when the product is put to uses that it is reasonable to expect a seller or distributor to foresee."), *with* § 2 cmt. d ("[T]he test is whether a reasonable alternative design would, at reasonable cost, have reduced the foreseeable risks of harm posed by the product.").

use of the product.[108] For example, one foreseeable use of a hammer could be driving a nail through a piece of wood.[109] Foreseeability of harm, in contrast, refers to the potential harms derived from the foreseeable uses of a product.[110] Thus, one foreseeable harm derived from using a hammer may be a broken finger from a missed swing.[111] This split-concept approach to foreseeability is important because while a manufacturer can take steps to limit the foreseeable uses of his product through various means,[112] foreseeable harms generally cannot be limited once foreseeable use is established: they are simply coextensive with the use of the product.[113] This contention is the basis of one of the issues posed by the doctrine's application to machine learning, discussed in the following section.

C. Core Difficulties of Applying the Restatement (Third)'s Design Defect Standards to Machine Learning

The main issues in applying the *Restatement (Third)'s* design defect standard arise from its application to machine learning and issues relating to foreseeability of risk and probabilistic harm. Part One addresses issues stemming from foreseeability of risk.[114] Part Two

108 *See* Eshbach v. W. T. Grant's & Co., 481 F.2d 940, 942–43 (3d Cir. 1973) ("[T]he proper limits of responsibility for the defendant-seller here is whether the 'use' to which the product was put was intended or foreseeable (objectively reasonable) by the defendant."); Hockler v. William Powell Co., 129 A.D.3d 463, 463 (N.Y. App. Div. 2015) (holding that plaintiff could not recover on a design defect claim because dismantling a product did not constitute "a reasonably foreseeable use of a product.").

109 Using the same hammer as a boomerang would not be a foreseeable use of the product because the normal and anticipatable uses of a hammer will generally only include those uses associated with construction.

110 *See Eshbach*, 481 F.2d at 943 ("[I]t is foreseeability as to the use of the product which establishes the limits of the seller's responsibility.").

111 For example, an injury sustained to the head stemming from the hammer being thrown into the air would not present a foreseeable harm, considering no foreseeable use is presented. Whether or not the manufacturer of the hammer would be held liable for a broken finger is not considered, as that question hinges upon the evidence of a reasonable alternative design ameliorating that foreseeable risk of harm.

112 For example, advertising can play a role in the determination of whether the plaintiff's use of a product is foreseeable. *See* Garrison v. Wm. H. Clark Mun. Equipment Inc., 241 A.D.2d 872, 873 (N.Y.S. 1997) (discussing how a brochure marketing the product expanded the foreseeable use of that product beyond that the manufacturer claimed).

113 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. m (AM. L. INST. 1999) ("Once the plaintiff establishes that the product was put to a reasonably foreseeable use, physical risks of injury are generally known or knowable by experts in the field.").

114 While not discussed in this Comment, note that the foreseeability issue identified in Part IV.C.1. poses similar problems to the *Restatement (Third)'s* warning defect standard for a similar set of reasons.

addresses the issue of probabilistic harm and its practical and policy-based implications on the reasonable alternative design requirement imposed by the risk-utility test.

1. Foreseeable Risk as a Dependent Element

As previously explained, for the purposes of determining whether a product has a design defect, a foreseeable risk of harm consideration occurs when "the product is put to uses that is reasonable to expect a seller or distributor to foresee."[115] The split between foreseeable use and foreseeable risk focuses the overall foreseeability inquiry on the foreseeability of use, going so far as to support a presumption that a risk of harm is foreseeable so long as the product is foreseeably used.[116] This means that analysis in products liability cases focuses on foreseeable use and not foreseeable risk, due to the presumption that foreseeable risk is derived from foreseeable use.[117] Normally, this is an inherently logical proposition, as the risk of harm is generally coextensive with use and a designer can usually anticipate and be expected to guard against risks associated with foreseeable use.[118]

Yet, when considering machine learning, this proposition is less sound due to computing concepts known as covariate shift and concept drift. Covariate shift occurs "when data fed into an algorithm during its use differs from the data that trained it" in some appreciable way.[119] A basic, harm-free example of this occurring in practice would be if "an imaging processing system" trained exclusively on laboratory conditions was "deployed to foreign geographic regions where light conditions differ."[120] In this case, the data used to train the machine learning system to perform its function will be starkly different from the

115 *Id.* § 2 cmt. m.

116 *See id.* § 2 cmt. m ("Product sellers and distributors are not required to foresee and take precautions against every conceivable mode of use and abuse to which their products might be put . . . In cases involving a claim of design defect in a mechanical product, foreseeability of risk is rarely an issue . . . Once the plaintiff establishes that the product was put to a reasonably foreseeable use, physical risks of injury are generally known."). Machine-learning based products are all mechanical in nature, as they themselves are machines operating for an intended function.

117 *Id.*

118 *See generally* Butts v. OMG, Inc., 612 F. App'x 260, 262–63 (6th Cir. 2015) (discussing the foreseeable risks of a product in the context of foreseeable use of that product).

119 Boris Babic et al., *When Machine Learning Goes Off the Rails*, HARV. BUS. REV., Feb. 2021.

120 Steffen Bickel et al., *Discriminative Learning Under Covariate Shift*, J. MACH. LEARNING RSCH. (2009).

real-world data, meaning that the system as a whole will likely not be able to function as intended as a result of the change in data.[121] To illustrate covariate shift in a non-technological manner, imagine a law student studying for an exam with an outline from a different professor: the concept may be similar in form, but the component pieces are so different that the answer may not be fully accurate.

In contrast, concept drift "describes unforeseeable changes in the underlying distribution of [real-world] data over time."[122] To make this simpler, a "machine learning [system developed] for stock trading" trained on market data derived from "a period of low market volatility and high economic growth" would experience concept drift if it were implemented during "a recession."[123] In that situation, the predictive function of the machine learning system would not be reliable because the algorithm's premise would not apply to the present market conditions.[124] Returning to the aforementioned ill-prepared law student, concept drift can be illustrated by imagining the student studying for a property exam by exclusively reading Blackstone on Property: the central idea may be the same, but most of the material has been superseded over time.

The issue in practice stemming from covariate shift and concept drift is that the risks of a machine learning product are not always coextensive with the use of the product. To illustrate, imagine a futuristic car with a self-driving feature that heavily incorporates machine learning to perform its self-driving function.[125] As a product subject to design defect standards, the self-driving car has a foreseeable use derived from its intended function—to be driven with the self-driving feature active.[126] If the car were to crash, harming an occupant or other individual, the crash would be a presumptively foreseeable risk of the use of the self-driving function.[127] But this is not always the case due to covariate shift and concept drift. Say, for example, the self-

121 *Id.*

122 Jie Lu et al., *Learning Under Concept Drift: A Review,* INST. ELEC. & ELEC. ENG'R (2018).

123 Babic et al., *supra* note 119.

124 Babic et al., *supra* note 119.

125 Gina Mantica, *Self-Taught, Self-Driving Cars? Like Babies Learning to Walk, Autonomous Vehicles Learn to Drive by Mimicking* Others, B.U.: THE BRINK BLOG (July 30, 2021), https://www.bu.edu/articles/2021/self-taught-self-driving-cars/ ("Self-driving cars are powered by machine learning algorithms that require vast amounts of driving data in order to function safely.").

126 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. m (AM. L. INST. 1999).

127 *See id.* ("Once the plaintiff establishes that the product was put to a reasonably foreseeable use, physical risks of injury are generally known.").

driving function failed because another driver near the car rapidly crossed three lanes of traffic.[128] If the implicit assumption made in the training data used to train the car's self-driving function was that other drivers would only ever unsafely cross two lanes of traffic, the machine learning system powering the self-driving function may experience covariate shift, resulting in an error that crashes the car.[129] The core problem is that while the ultimate risk in such an example, the car crashing, is absolutely foreseeable to the designer, the instrumentalities by which that risk arises are not.

The instances leading to potential covariate shift and concept drift are functionally endless. This concept, the immeasurability of the points of failure, is generally known as the black box of machine learning.[130] More specifically, the black box of machine learning refers to the idea that "no human can understand how . . . variables are jointly related to each other to reach a final prediction," essentially that even the designer of a machine learning system does not know the exact steps the system takes to reach its conclusion.[131] One implication of covariate shift and concept drift is that it may not be readily apparent to a designer exactly where a point of failure lies in the machine learning system due to the black box.[132] Thus, even minor changes between the training dataset and real-world data leading to covariate shift can be a monstrous endeavor to diagnose and fix due to the black box generally obscuring the inner workings of the system.[133]

The broader issue, for the designer at least, is that while covariate shift and concept drift are foreseeable in occurrence, their effects and associated risks are frequently not. A designer can make a relatively safe bet that the machine learning system designed will at some point

128 *See generally Distribution-Shift—The Hidden Reason Self-Driving Cars Aren't Safe Yet*, MEDIUM (Apr. 14, 2020), https://medium.com/@nuronlabs/distribution-shift-the-hidden-reason-self-driving-cars-arent-safe-yet-b07bfe3ae800 (discussing possible examples of covariate shift self-driving cars may experience).

129 *See* Babic, *supra* note 119 (providing the definition and examples of covariate shift in machine learning).

130 Cynthia Rudin & Joanna Radin, *Why Are We Using Black Box Models in AI When We Don't Need To? A Lesson From an Explainable AI Competition*, HARV. DATA SCI. REV. (Nov. 22, 2019).

131 *Id.*

132 *See id.* (stating that "even if one has a list of input variables, black box predictive models can be such complicated functions of the variables" that a human designer cannot identify how the system reached its conclusion).

133 *Id.* The same would be true of a change in real-world conditions that leads to the machine learning system experiencing concept drift.

experience either, or both, covariate shift and concept drift,[134] but the same cannot be said of what harms and risks of harm will result from this occurring. However, liability for the designer does not hinge on whether they make the astute observation that covariate shift will one day occur; liability is imposed when the ultimate harm stemming from covariate or concept drift is foreseeable.[135] Going back to the prior example, the question of liability for the designer of the self-driving car does not turn on whether he has correctly foreseen that covariate shift or concept drift will one day occur; liability is imposed when it occurs and causes harm, without regard for the circumstances that brought it about.[136]

The core issue then emerges: a plaintiff may experience a foreseeable harm from their own or another's use of the machine learning system in a foreseeable manner by means that are unforeseeable to the designer "at the time of sale."[137] To make this more concrete, recall the prior example of the futuristic self-driving car powered by a machine learning system.[138] The foreseeable use of this product is the use of the self-driving feature to drive the car,[139] and a foreseeable risk of harm stemming from the foreseeable use of the self-driving feature is the risk that the car may crash.[140] If the self-driving feature of the car malfunctions, leading to a crash, the plaintiff presumptively satisfies the foreseeability inquiry: the product has been used in a foreseeable manner and caused a foreseeable harm.[141] This result is massively overinclusive towards the satisfaction of the foreseeability requirement because it does not account for the near-

134 *See Data Distribution Shifts and Monitoring*, CHIP HUYEN (Feb 7, 2022), https://huyenchip.com/2022/02/07/data-distribution-shifts-and-monitoring.html ("Data shifts happen all the time, suddenly, gradually, or seasonally.").

135 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(b) (AM. L. INST. 1999).

136 Therefore, the question of whether or not a car swerving across three lanes of traffic is foreseeable to the designer is irrelevant under the current design defect standard. Liability turns on whether the ultimate harm experienced by the driver is a foreseeable risk of the foreseeable use of using the autonomous driving mode. *Id.*

137 *Id.* at § 2(b) (AM. L. INST. 1999).

138 *See* Mantica, *supra* note 125.

139 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. m (AM. L. INST. 1999).

140 *Id.* ("Once the plaintiff establishes that the product was put to a reasonably foreseeable use, physical risks of injury are generally known or knowable by experts in the field."). This is an example of a generally known risk, in that it is fairly common knowledge that driving a car carries a risk that the car will crash, machine learning system driving or not.

141 *See id.* (stating that design defect liability is "impose[d] only when the product is put to foreseeable uses."); *Id.* at § 2(b) (stating that a product "is defective in design when the foreseeable risks of harm posed by the product" could have been lessened).

infinite number of variables leading to the foreseeable harm stemming from the foreseeable use. While some of these variables may be within the designer's control,[142] many are not and are, in fact, caused by extrinsic phenomena that produce covariate shift and concept drift.

Thus, this discounting of the interim step between foreseeable use and foreseeable harm does not account for the technological realities of machine learning. It is simply not possible for a designer to program around every conceivable edge case that may lead to a foreseeable harm.[143] But if the designer cannot do this, then their liability for harm is essentially endless as well, so long as the foreseeability inquiry mirrors that of the example provided. The use and risk of harm are not what is unforeseeable to the designer, but the instrumentalities of that risk of harm are, thereby posing a unique problem for the designer of the machine learning system.[144]

The question then becomes, what is a designer supposed to do to counteract this problem? Generally, broadening training data to the extent that every conceivable edge case or cause of harm is incorporated is not technologically possible.[145] The implicit and necessary assumption that a designer must make at the outset of developing a machine learning system is that a certain subset of edge cases will not be designed around or accounted for in any meaningful way.[146] Realities concerning machine learning function conflict with the law in this regard because there are effectively no means by which a designer can limit liability other than to simply not develop or incorporate the machine learning system.

This result does not mesh with the stated policy goals of the *Restatement (Third)*. Clearly stated is that a designer need not "take

142 For example, the self-driving car may simply fail as a result of poor programming, which is in the control of the designer.

143 *See* Jason Withrow, *Edge Cases: A Persistent Dilemma*, STOUT SYS., https://www.stoutsystems.com/edges-cases-a-persistent-dilemma/ (discussing the impossibility of eliminating edge cases in machine learning design and practice) [perma.cc/V2P3-Y4BK].

144 A loosely comparable situation may be found in the designers of prescription medication. For such products, unforeseeable risks of harm such as specific individual's reactions to medication may be unforeseeable at the time of design. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. m (AM. L. INST. 1999). However, for machine learning systems, the difference lies in the fact that the ultimate risk of harm is generally observable, but the instrumentalities of that harm are unforeseeable at the time of sale.

145 *See* Amal Joby, *What is Training Data? How it's Used in Machine Learning*, G2 (Jul. 30, 2021), https://learn.g2.com/training-data (emphasizing that training data must be "relevant"); Withrow, *supra* note 143. (discussing the impossibility of eliminating edge cases in machine learning design and practice).

146 Withrow, *supra* note 143.

precautions against every conceivable mode of use and abuse" when designing a product.[147] However, by disregarding the instrumentality of the harm that a machine learning system causes, the *Restatement (Third)* essentially asks a designer to do exactly that. To avoid liability, the designer must essentially include every possible edge case that may be reasonably expected to lead to a foreseeable risk during foreseeable use. The post hoc conclusion that the designer should have included a system reaction to a specific edge case is simply a new way of saying that every mode of use must be addressed in the design. While the designer of every product picks and chooses risks of harm to mitigate to a degree, machine learning systems are not afforded the same balance of risk and benefit that other products have due to the lack of focus on the instrumentalities of risk.

2. Probabilistic Harm and Reasonable Alternative Designs

Generally, before applying the risk-utility test, the plaintiff in a design defect case must present evidence that establishes a reasonable alternative design.[148] A reasonable alternative design is one that "would, at reasonable cost, have reduced the foreseeable risks of harm posed by the product" and that "could have been practically adopted at time of sale" without being overly costly.[149]

The difficulty in applying this requirement to products incorporating machine learning derives from the systems' functionality. At its core, a machine learning system is a system designed to produce outputs based on probabilistic evidence.[150] Probability theory is a foundational concept in machine learning because probability informs the system on the outcomes likely to occur when a specific course of action is taken or when a certain combination of variables are present.[151] However, probability is not an exact science; what is likely to occur is not certain to occur. For example, consider calculating the probability of a coin flip landing on heads. The laws of probability state that either option, heads or tails, has a 50 percent probability of occurring, but flipping a coin one hundred times rarely leads to an even

147 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. m (AM. L. INST. 1999).

148 *Id.* at § 2(b). There are situations in which this bar need not be met, but the general rule is that such evidence must be presented. See *id.* at § 3 cmt. b (discussing situations in which a reasonable alternative design need not be shown by a plaintiff).

149 *Id.* § 2 cmt. d (AM. L. INST. 1999).

150 *See generally* Christopher M. Bishop, PATTERN RECOGNITION AND MACHINE LEARNING (M. Jordan et al. eds., 2006) (discussing machine learning algorithms as pattern recognition models based on probability theory).

151 *Id.*

distribution of heads and tails.[152] This is because the most probable outcome, an even distribution, is not the only outcome that can occur—it is simply the most probable or likely outcome.[153]

When applying the realities of probability theory to machine learning products, the mismatch between probable and actual results has the potential to cause harm. Take, for example, a fraud detection system that uses a machine learning algorithm to flag fraudulent transactions used by a financial institution.[154] If the system designates a specific transaction as fraudulent with a 99 percent probability, and the account is flagged, the corollary assumption is that there is a 1 percent probability that the transaction is not fraudulent. If the transaction falls into that 1 percent category, the system has effectively harmed the individual by potentially affecting their credit, among other repercussions.[155] However, the system has done exactly what it was designed to: identify the most probable output and act accordingly, even though that output had a small chance of causing harm.

What, then, is the alternative design that ought to be proposed when an individual is harmed based on a statistical improbability? In the above example, any proposed alternative design would effectively need to eliminate the chance for a statistical improbability to occur for the plaintiff to be able to show that the design "would have reduced or prevented injury to" them, as the designer would need to disallow the system from making decisions based on anything less than 100 percent probability, so long as that alternative was of reasonable cost and availability.[156] But such a design is antithetical to the purposes behind using a machine learning system in the first place: why design a system

152 John Walker, *Introduction to Probability and Statistics*, THE RETROPSYCHOKINESIS PROJECT, https://www.fourmilab.ch/rpkp/experiments/statistics.html.

153 *Id.*

154 *See generally* Kaushik Choudhury, *Real-Time Fraud Detection with Machine Learning*, MEDIUM: TOWARDS DATA SCI. (Sept. 2, 2020), https://towardsdatascience.com/real-time-fraud-detection-with-machine-learning-485fa502087e (discussing the use of machine learning algorithms in real-time fraud detection for financial institutions); Florian Tanant, *Fraud Detection with Machine Learning & AI*, SEON, https://seon.io/resources/fraud-detection-with-machine-learning/ (discussing the use of machine learning in financial fraud detection and the benefits of its use).

155 Discussion of the merits of such a claim are beyond the scope of this Comment, but depending on the facts of the case, the consumer may have a claim under The Fair Credit Reporting Act, 15 U.S.C. § 1681(a)(1)–(2).

156 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. d (AM. L. INST. 1999).

to analyze patterns in data to reach a conclusion when one can simply design a system to reach a conclusion based on preformulated inputs?[157]

In this way, the proposed alternative design to the fraud detection system that eliminates statistical improbabilities looks more akin to the assertion that the product design is manifestly unreasonable.[158] A manifestly unreasonable design is one that has low social utility coupled with a high risk of harm, for which a plaintiff need not present evidence of a reasonable alternative design.[159] By asserting that the only way to make a machine learning system reasonable is to eliminate statistically improbable outcomes, as would be required of the fraud detection system, the plaintiff is effectively stating that the design is manifestly unreasonable because any evidence of an alternative design would not be the same product, conceptually or otherwise.[160] Instead, the alternative design in such a scenario would effectively be an entirely different system that operates without the central feature of the design—the probability-based decision-making function. However, unlike the average manifestly unreasonable design, machine learning systems like the fraud detection system have social and societal utility when considered in the broader context of their use. A plaintiff succeeding on a claim that such a system is, in effect, manifestly unreasonable detracts from the widespread utility of such systems and does not further the ends sought to be furthered by design defect doctrine.[161] Machine learning, while not perfect, generally provides societal benefits in the form of risk-reduction through elimination of human error and enhanced safety through automation;[162] holding machine learning products to be manifestly unreasonable detracts from

157 Once predictive nature of the system is eliminated, the system operates as a rules-based system. This is because instead of analyzing data to reach a conclusion, the system is simply looking for data—defined by rules—that confirms a conclusion. For a discussion of rules-based systems, see ROUTLEDGE, *supra* note 47, at 454.

158 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. e (AM. L. INST. 1999). The *Restatement (Third)* does not expressly endorse the idea of manifestly unreasonable design but it does discuss the doctrine in relation to potential analysis of a products liability claim.

159 *Id.*

160 *Id.*

161 *Id.* § 2 cmt. e ("The court would declare the product design to be defective and not reasonably safe because the extremely high degree of danger posed by its use or consumption so substantially outweighs its negligible social utility that no rational, reasonable person, fully aware of the relevant facts would choose to use . . . the product.").

162 *The Top 5 Benefits of Artificial Intelligence*, DEFINED AI (Nov. 23, 2020), https://www.defined.ai/blog/the-top-5-reasons-to-be-grateful-for-ai/?WPACFallback=1&WPACRandom=1679675641431.

benefits, as designers seeking to use the technology are faced with significant litigation risk.

Additionally, the policy behind design defects is not entirely applicable to machine learning systems causing probability-based harm. A finding that a product is defectively designed essentially states that "every unit in the same product line is potentially defective," meaning that the product's design should be changed to avoid future liability.[163] Therefore, designers are encouraged to use a reasonable alternative design or an equivalent instead of the current design for reasons of limiting liability.[164] But, the mere occurrence of a statistical improbability actually occurring does not allow for the inference that every identical system in use is defective. Again, the products function as intended, even when they make a probability-based decision that causes harm. Returning to the prior fraud detection system example, an exact replica of the system deployed elsewhere looking at a substantially similar transaction could make the exact same decision as the system held defective. If this new transaction is fraudulent, however, then the supposed defect found in the original system does not apply to the identical system elsewhere. In any case, the important point is that the harm experienced in such a case is, in all actuality, simply a manifestation of the inherent problem with making probability-based decisions and not a failing of the system itself.

The broad takeaway from this discussion is that, at least for some types of machine learning systems, the current conception of what constitutes a defect can be troublesome for designers. Not only does the reasonable alternative design requirement not operate as intended for these systems, but the policy implications underlying holding them defective may also not comport with technical realities.

V. POSSIBLE APPROACHES TO RECTIFYING THE ISSUES POSED BY MACHINE LEARNING

Presenting an answer to the previously identified issues is as complex as the question itself because any conclusion drawn will necessarily reflect the value one places on machine learning as a technology and the relative proportion of harm one tolerates to advance it. Recognizing this value-based reasoning, this Comment provides

163 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 1 cmt. a (AM. L. INST. 1999).

164 *Id.* at § 2 cmt. a ("The emphasis is on creating incentives for manufacturers to achieve optimal levels of safety in designing and marketing products.").

three alternative solutions to the issues discussed, each reflecting a different attitude toward machine learning and its advancement.[165]

This section proceeds in three parts. First, Part A provides a solution reflecting the sentiment that machine learning's inclusion in a product represents an inherent defect, thereby placing the lowest possible value on machine learning as a technology. Next, Part B suggests a middle-ground reformational approach to current product liability doctrine to solve the previously identified issues, thus placing an even value on machine learning and its relative propensity to cause harm. Finally, Part C offers an optimistic cascade theory based on the idea that machine learning's value is higher than that of its ability to cause harm.[166]

A. Machine Learning as a Defect

One may reasonably conclude from the prior discussion that the inclusion of a machine learning element in a product represents an inherent defect in that product for the purposes of product liability.[167] Justification for this theory may be premised on the idea that the foreseeability issues presented by machine learning and the capacity for probability-based harms are simply too great in light of the language of the *Restatement (Third)* and its related policy justifications.[168]

The basic premise of such an argument rests on the idea that there is always a reasonable alternative design available for a product incorporating machine learning: a version of the product not incorporating machine learning. Essentially, because the designer cannot foresee all instrumentalities leading to foreseeable risks of harm and the machine learning system always carries a risk of probabilistic

165 In short, the solutions proposed by this Comment harken back to the most frustrating answer one can give to a legal question: it depends. Susan Landrum, *The Most Frustrating Phrase in Law School: "It Depends"*, L. SCH. SUCCESS (Aug. 14, 2014, 8:00 AM), https://lawschoolacademicsuccess.com/2014/08/14/the-most-frustrating-phrase-in-law-school-it-depends/.

166 While somewhat of a misnomer, what this Comment coins as cascade theory may also be accurately described as the adoption of machine learning strategies by industry laggards under the Rogers adoption curve model. *See* Milo Miszewski, *Technology Adoption Curve–Everything That You Need to Know*, MDEVELOPERS (May 27, 2021), https://mdevelopers.com/blog/technology-adoption-curve-everything-that-you-need-to-know (explaining the Rogers technology adoption curve in relation to consumer reactions to new technologies).

167 Courts have taken similar positions in other contexts, stating that some conditions, "even if resulting from the design of the products, are defective." *See, e.g.*, Phipps v. General Motors Corp., 363 A.2d 955, 959 (Md. 1976).

168 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 cmt. a (AM. L. INST. 1999) (emphasizing safety as the driving force behind the creation of strict products liability).

harm, then machine learning itself functions in a defective manner. Until these issues with machine learning are solved, then machine learning is defective because a system designed without machine learning will always "reduce[] or avoid[]" risks of harm.[169]

Thus, the solution proposed under this view functions as a simple yes or no question to determine if a machine learning system was involved in the harm a plaintiff experiences. If that question is answered in the affirmative, then the product will always contain a defect due to the easy evidence of a reasonable alternative design.[170] Liability for the designer, therefore, is somewhat presumed in such a situation, as the presumed nature of the defect and reasonable alternative design simply leaves no room for litigation on the issue.[171] Thus, strict product liability becomes even stricter for the designers of machine learning products.

Obviously, this perspective does not place a high value on machine learning technology, especially in instances where it has a possibility of causing harm. Thus, consumer safety and adherence to existing doctrine and its related policy goals outweigh any prudential considerations of the usefulness of machine learning in consumer products: if the technology cannot overcome the reasonable alternative design requirement imposed due to its nature, it simply must be classified as defective until the technology develops further.

B. Reformational Approach

If one values machine learning technology while recognizing that current design defect standards may be retooled to compensate for the unique issues posed by the technology, then a reformational approach emerges. There is, however, a fundamental problem with proposing exacting changes to the law when dealing with rapidly developing technology: the technology has the chance to outpace any changes made in the law to accommodate it.[172] Thus, the recommendations in this section should be viewed as benchmarks in that they operate as a

169 *Id.* at § 2(b).

170 *See generally* Phipps, 363 A.2d at 959 (stating that products containing an inherent defect do not require "weighing and balancing the various factors involved.").

171 *See* Cavanaugh v. Stryker Corp., 308 So. 3d 149, 154 (Fla. Dist. Ct. App. 2020) (implying that proof of a reasonable alternative design plays a "central role" in design defect cases).

172 *See* Julia Griffith, *A Losing Game: The Law is Struggling to Keep Up with Technology*, SUFFOLK J. OF HIGH TECH. L. BLOG (Apr. 12, 2019), https://sites.suffolk.edu/jhtl/2019/04/12/a-losing-game-the-law-is-struggling-to-keep-up-with-technology/ (providing examples of legal standards and rules that have not kept pace with technology).

standard for measuring whether any future changes in law will adequately compensate for the challenges machine learning technologies pose.

1. Foreseeability of the Instrumentality of Risk as an Independent Standard

To adequately consider the unique nature of machine learning, any future design defect test must incorporate the foreseeability of the instrumentalities of foreseeable risks. As explained, the current standard's exclusive focus on foreseeable use and foreseeable risk leads to unlimited liability for designers of machine learning systems due to the unlimited range of instrumentalities leading to foreseeable risks.[173] Including this extra component will primarily serve to limit the designer's liability. However, it will not go so far as to immunize them for those instrumentalities of harm that are truly foreseeable.

To illustrate, return to the example of the self-driving car crashing because of the machine learning system experiencing covariate shift or concept drift. If the cause of the covariate shift or concept drift was foreseeable, for example, the machine learning system could not recognize the side of a truck as another vehicle,[174] then the vehicle designer would not be absolved of liability. But if the cause were not foreseeable, maybe due to a sudden swarm of cicadas covering the vehicle,[175] the vehicle designers would have an avenue to immunize themselves of liability for the harm.

The key point is that the insertion of this independent factor leads to more equitable results in applying design defect standards. Plaintiffs will still be able to hold the designers of truly defective machine learning systems responsible for defects, as satisfaction of this additional element will not be a high bar to cross.[176] At the same time, however,

173 *See supra* Part IV.C.1.

174 Timothy B. Lee, *"I Was Just Shaking"—New Documents Reveal Details of Fatal Tesla Crash*, ARS TECHNICA (Feb. 15, 2020, 9:00 AM), https://arstechnica.com/cars/2020/02/i-was-just-shaking-new-documents-reveal-details-of-fatal-tesla-crash/ (speculating that "[t]he machine learning algorithms that underpin [the self-driving system] have only been trained to recognize the rear of other vehicles, not profiles or other aspects.").

175 While somewhat of a ridiculous example, it proves the ultimate point of the necessity of this independent component of liability: currently, if the car crashed as a result of this example occurring, both the use and risk of harm would be foreseeable to the designer at the time of sale. *See supra* notes 136–137 and accompanying text.

176 The *Restatement (Third)* acknowledges this point broadly, stating that "in cases involving a claim of design defect in a mechanical product, foreseeability of risk is rarely an issue as a practical matter." RESTATEMENT (THIRD) OF TORTS: PRODS LIAB. § 2 cmt. a (AM. L. INST. 1999).

designers can rest easier knowing that their liability for such products is less than the whole of human experience.

This approach also meshes well with the existing policy goals of the R*estatement (Third)*. As explained previously, the current standard effectively forces a designer to design for every possible reality due to the lack of focus on the cause of the risks of harm, which does not keep with the policy goals of design defect liability in that it creates an overbroad category of liability for designers of machine learning products.[177] By not discounting the interim step between foreseeable use and foreseeable risks of harm, design defect liability will be more in line with this policy goal because designers of machine learning products will only have to design for truly foreseeable scenarios and not every conceivable instrumentality of harm.

2. Probabilistic Harm Proofs

Any future design defect test must adequately account for the chance of probabilistic harm. As explained, machine learning, being at its core a technology used to make probability-based predictions, has the potential to cause harm based on individuals being on the wrong side of probability.[178] Solving this issue is more complicated than it seems at first glance, considering the functionality of machine learning technology.

The obvious solution is to prescribe, legislatively or otherwise, a probability floor at which a machine learning system capable of causing harm is allowed to decide. For example, it may be prescribed that the credit fraud detection system from earlier is only allowed to flag an account when it is 99% sure of the transaction's fraudulent nature.[179] However, prescribing such a bar may hinder machine learning technology to the point of irrelevancy, as the whole point of machine learning is to make predictions on unclear data: if minimum confidence is required, a rules-based system may actually perform better.[180]

Thus, the better solution is to impose a requirement that the difference in probability between the machine learning system and the reasonable alternative design must represent a material change in

177 *See id.* at § 2 cmt. m (discussing the policy behind the inclusion of the foreseeable use and foreseeable risk elements of design defect liability).

178 *See supra* Part IV.C.2.

179 *See supra* note 154 and accompanying text.

180 *See supra* note 154 and accompanying text.

statistical likelihood.[181] Thus, alternative design proofs that merely show a slight statistical deviation will not be considered adequate in the context of machine learning. While this somewhat overwrites the text of the rule in this context,[182] the alternative provides too easy a route to liability for a plaintiff attempting to prove a reasonable alternative design through minor statistical deviations.

Thus, to illustrate, suppose the previously mentioned fraud detection system was allowed to flag an account with only a 75% chance of fraud. If a plaintiff were to present proof that a reasonable alternative design existed in the form of a system only allowed to flag transactions with a 67% probability of fraud, this 1% difference would likely not be significant enough to satisfy the materiality requirement. However, proof of a system only allowed to flag transactions with a 99% probability of fraud likely would. In either case, the reasonable alternative design would still need to satisfy the other requirements for proving a reasonable alternative design, namely that the design existed at the time the plaintiff was harmed.[183]

C. Cascade Theory of Machine Learning

The corollary position to the total classification of machine learning as a defect is the idea that, for some products, lacking a machine learning element represents an inherent product defect due to the benefits that machine learning provides. The premise of this argument rests on the idea that automation traditionally enhances safety,[184] and therefore, automating away human input from some classes of products will reduce the possibility of human error.

This position may hold water in the world of self-driving vehicles. High levels of automation promise to "remove the driver from the chain of events that can lead to a crash," thereby eliminating human error

181 In this context, material meaning "significant; essential." *Material*, BLACK'S LAW DICTIONARY (11th ed. 2019).

182 Essentially, such a change would disregard the part of the rule only requiring a reasonable alternative design to "reduce[] or avoid[]" risks of harm to the consumer. RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2(b) (AM. L. INST. 1999).

183 *Id.* § 2(b) cmt. d.

184 For example, automation in manufacturing reduces the risk of accidents and workplace injuries. Erika Strand & Peter Stipan, *Using Automation Technology to Improve Facility Safety*, EHS TODAY (Dec. 18, 2020), https://www.ehstoday.com/safety-technology/article/21150776/using-automation-technology-to-improve-facility-safety; Jim Vinoski, *What's Automation Ever Done for Us? Okay, There is the Improvement in Worker Safety*, FORBES (Dec. 7, 2018, 1:30 PM), https://www.forbes.com/sites/jimvinoski/2018/12/07/whats-automation-ever-done-for-us-okay-there-is-the-improvement-in-worker-safety/?sh=3aa02c91771e.

behind the wheel.[185] Currently, human error causes about 94 percent of all auto accidents.[186] Therefore, if self-driving technology were to be greatly improved, the brunt of traffic accidents may be eliminated. While the precise number of accidents that will be prevented is in contention,[187] with some conservative estimates claiming that only about a third of accidents will be prevented, the broad point is that driver safety will be significantly enhanced. While achieving this reduction may be a way away,[188] there is little debate on whether it is capable of being achieved through the use of machine learning technology.[189]

This same line of reasoning may extend elsewhere. For example, machine learning software used to detect and prevent warehouse injuries can reduce workplace accidents and injuries by eighty percent.[190] Medical algorithms used to evaluate chest X-rays produce better results than expert radiologists in diagnosing specific diseases.[191] Predictive policing, accomplished through machine learning, has shown potential in reducing crime.[192] In summation, there are many places

185 *Automated Vehicles for Safety*, NHTSA, https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety (last visited Jan. 14, 2024).

186 NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., U.S. DEP'T OF TRANSP., DOT-HS-812-115, CRITICAL REASONS FOR CRASHES INVESTIGATED IN THE NATIONAL MOTOR VEHICLE CRASH CAUSATION SURVEY (2015), https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115. The remaining six percent may be attributed to vehicle component degradation, environmental hazards, and other unknown critical reasons.

187 Jaime Ramos, *Autonomous Vehicles and Accidents: Are They Safer Than Vehicles Operated by Drivers*, TOMORROW.CITY (June 22, 2022), https://tomorrow.city/a/self-driving-car-accident-rate (stating that estimates on how many accidents would be eliminated vary from "the complete eradication of that 94%, to more pessimistic reports (by insurance companies) that calculate a reduction of around 35%.").

188 *Automated Vehicles for Safety*, NHTSA https://www.nhtsa.gov/technology-innovation/automated-vehicles-safety (predicting 2025 and beyond as the earliest that fully autonomous driving may be achieved) (last visited Jan. 17, 2024).

189 *Id.* (predicting that "advantages of [machine learning self-driving] technology could be far-reaching.").

190 *Applying Machine Learning to Keep Employees Safe and Save Lives*, VENTURE BEAT (Sep. 29, 2020, 6:50 AM), https://venturebeat.com/ai/applying-machine-learning-to-keep-employees-safe-and-save-lives/ (discussing the machine learning platform Warny, which leads to, "on average, . . . an 80% drop in incidents,").

191 Taylor Kubota, *Stanford Algorithm Can Diagnose Pneumonia Better Than Radiologists*, STANFORD NEWS (Nov. 15, 2017), https://news.stanford.edu/2017/11/15/algorithm-outperforms-radiologists-diagnosing-pneumonia/ (discussing CheXNet, a machine learning algorithm designed to diagnose fourteen types of medical conditions based on analysis of chest X-rays).

192 Matt Stroud, *Official Police Business: Does Predictive Policing Actually Work?*, VERGE (May 4, 2016, 12:04 PM), https://www.theverge.com/2016/5/4/11583204/official-police-business-predictive-policing-paper (explaining that PredPol, a predictive policing software, "can lead to a 7.4 percent reduction in 'crime volume.'"). *Contra*

where the use of machine learning technology enhances the safety of consumers and individuals, even considering the legal difficulties created by the technology.

Similar to machine learning as a defect theory, this position would simply ask at the outset if an alternative with a machine learning element that enhances safety was available to the product designer.[193] If there was, then the lack of machine learning represents an inherent defect in the product in every instance.[194] In such cases, liability is again nearly presumed due to the simple proof of an alternative design for the product in question.[195]

This approach places an extremely high value on machine learning as a technology, to the end that its use will become nearly ubiquitous over time. While the same set of issues posed previously might still exist, the current holes in the standards might be smoothed out due to increased litigation on the topics and enhanced awareness of the unique issues. In essence, this position disregards existing law in favor of allowing rapid progress in machine learning and realizing the expected safety benefits of the near-total use of machine learning technology.

VI. CONCLUSION

In summary, it is possible to remedy the current inadequacies of design defect liability when applied to machine learning in several ways, depending on the value one places on the technology. In any case, the issues identified by this Comment will not go away on their own and are likely to expand as more and more products incorporating machine learning enter the market for consumer use. Ultimately, it will be necessary to reformat and retool the standards used to ensure that the law does not become entirely outdated in the face of ever-changing and continually advancing technology.

Andrew G. Ferguson, *Policing Predictive Policing*, 94 WASH. U. L. REV. 1109, 1117 (2017) (broadly criticizing predictive policing's effectiveness).

193 RESTATEMENT (THIRD) OF TORTS: PRODS. LIAB. § 2 (AM. L. INST. 1999).

194 *See supra* notes 163, 166.

195 *See supra* note 167.