Woodrow Barfield*

Liability for autonomous and artificially intelligent robots

https://doi.org/10.1515/pjbr-2018-0018
Received February 10, 2018; accepted July 12, 2018

Abstract: In the backdrop of increasingly intelligent machines, important issues of law have been raised by the use of robots that operate autonomous from human supervisory control. In particular, when systems operating with autonomous robot’s damage property or injure humans, it may be difficult to determine who’s at fault and therefore liable under current legal schemes. This paper reviews product liability and negligence tort law which may be used to allocate liability for robots that damage property or cause injury. Further, the paper concludes with a discussion of different approaches to allocating liability in an age of increasingly intelligent and autonomous robots directed by sophisticated algorithms, analytical, and computational techniques.

Keywords: robot, negligence, products liability, artificial intelligence, autonomy, algorithm

1 Introduction: law, robots, and liability

With advances in artificial intelligence (AI), and particularly with machine learning techniques which allow robots to learn from experience and to solve problems using algorithms and sophisticated analytical techniques, robots are becoming more and more independent from human supervisory control. Unlike many of the non-autonomous robots operating in our homes and factories, autonomous robots are able to take a high-level goal and determine how to achieve it. Often with little or no human supervision at all. In fact, to some extent robots operating with machine learning techniques can learn how to solve problems rather than having to be pre-programed to perform each individual task [1]. Given techniques from machine learning, there is the possibility that robots may develop novel solutions to problems that take humans completely out of the decision-making loops of the system. Further, with robots that have the ability to learn and to use novel solutions to solve problems, it may be difficult to know in advance whether this class of robots will damage property or cause injuries. In this case, determining who is at fault or liable for any damages resulting from the robot’s actions may be difficult or even impossible under current legal schemes [2, 3]. On this point, Andreas Matthias has discussed how “autonomous learning machines” may create situations where the manufacturer/operator cannot be held morally or legally liable for the machine’s behavior due to the inability to accurately predict the machine’s future actions [4].

This paper focuses on legal issues associated with robots that take in sensor information about the environment and using machine learning techniques, learn about the world, navigate, and make decisions autonomously. Robots with autonomy are generally able to work for some period of time independently of human intervention, navigate independently without human assistance, and avoid situations which may be harmful to the robot, people, or property. These characteristics of robots may lead to self-maintenance, independent navigation throughout the environment, and the learning required to perform tasks autonomously. But with such robots, one problem for the law is that robots with machine learning, may not learn or reason like humans do, and that can make their outputs difficult to predict, explain, and analyze under current legal schemes [1, 2].

As part of the discussion on issues of liability for robots, this paper briefly reviews basic concepts of tort and contract law which apply to robots gaining intelligence and autonomy. Additionally, some case law involving non-autonomous robots is presented, with the goal to inform the reader of the types of disputes that have occurred involving robots and how the courts have analyzed such cases. Since previously decided cases serve as precedence for future legal disputes, the body of “robotic law” which exists now reflects a valuable source of law for a future in which autonomous robots will be involved in disputes that will find their way to court. Considering autonomous

*Corresponding Author: Woodrow Barfield:
Professor Emeritus, University of Washington, Seattle, USA,
E-mail: Jbar5377@gmail.com

Unauthenticated Download Date | 7/6/19 8:20 AM
robots, regulation and liability can be thought of as two sides of the same safety/public welfare coin. Regulation is about ensuring that intelligent systems are as safe as possible while still spurring innovation; liability is about establishing who is to blame with system failures—or, more accurately, who society can extract legal redress from—when something goes wrong.

According to legal scholar Andrew Tutt, “Algorithms that engage in machine learning differ fundamentally from other [standard] algorithms” [1]; for example, he comments that they are more complex, opaque, and less predictable. In addition, Zachary Lipton provides a more in-depth discussion of model interpretability specifically for supervised learning and comments that interpretability of algorithms is not monolithic and reflects several distinct ideas [5]. And according to judge Curtis Karnow, robots with machine learning are able to determine their own means to accomplish goals and their behaviors may not be predictable by either the operator-owner or by the original programmers [6]. With machine learning techniques, the general idea that the control logic operating a robot could be a causative factor in an accident has relevance for determining liability for autonomous systems; this basic concept was expressed in the context of human decision makers in the early case of Greenman v. Yuba Power Prod. Inc., in which the court stated “[a] defect may emerge from the mind of the designer as well as from the hands of the workman” [7]. With current machine learning techniques, so too can liability for harm to humans or property damage arise from the “mind of the robot,” yet no current legal scheme would hold the robot liable for its actions, no matter how autonomous the machine.

Within the field of robotics, there has been a transition in the control of robots operating with predetermined logic and rules, and with little capacity to learn; to robots expressing some level of autonomy; and finally, to autonomous robots with the ability to learn (at least for some tasks). This transition has resulted in challenges to the law which deals directly with liability for injury or property damage. Thus, as machine learning techniques control more of the robot’s behavior, the issue of determining who is liable for damages that may occur is an important question for courts to consider. And while some intelligent machines may be a specialist in a narrow domain, such as the supercomputer Deep Blue in chess [8], machines equipped with machine learning techniques can learn in a wider problem-solving domain and use novel solutions to solve real-world problems; thus, offering distinct challenges to courts. For example, with techniques such as reinforcement learning, robots are able to perform physical tasks and respond to the environment without human supervision. Under this scenario, robots may perform tasks with solutions that were unpredictable in advance and do so using solutions that are often novel and even unknown to humans. When such systems are involved in transactions which lead to injury or property damage, the issue of determining who is liable becomes difficult because no human may be identified as a causative factor in the chain of events leading to damages.

2 Robots with autonomy

Briefly, a robot that can act autonomously has the capability to learn from experience and thus to adopt to the demands of the environment. Such capabilities are creating a class of robots which are able to generalize behavior from one situation to another and that are beginning to interact in an intelligent way with everyday objects [6]. “Intelligent robots” may not need input from programmers for all tasks, instead, algorithms will be provided with some basic tools for problem solving and then will be left to construct for themselves their own tools to solve intermediate problems, on the way to solving abstract goals. Of course, under this scenario, determining who to hold liable if a robot acting autonomously, damages property or results in injuries, will be challenging.

Using machine learning techniques, robots can learn tasks which are beyond the capabilities of industrial and other types of robots that operate with a predetermined set of instructions [2, 3, 9]. In contrast to these robots, various machine learning approaches use algorithms that permits software to learn— to train itself to perform tasks like speech and image recognition. Since a robot is controlled by the algorithms directing its actions, for example, in the case of neural networks, the layers of the neural network where learning occurs, algorithms could be a contributing factor to injuries or property damage. In my view, it is within the machine learning software and algorithms controlling a robot, that the courts should focus in decisions that involve determining liability. As robots learn to solve problems using solutions unknown to a human operator and increase their independence in decision making and mobility, they are gaining the type of skills and abilities that could lead to actions that are wholly unpredictable, if so, who is at fault if damages to property or people occur, especially if no human was aware of the robot’s activities or was knowledgeable of the workings of the algorithms controlling the robot’s behavior?

As an example of the difficulty determining liability for intelligent robots under established law, consider Bax-
ter, a two-armed industrial robot designed to perform repeatable tasks. To increase the general intelligence and autonomy of Baxter, Pinto and Gupta of Carnegie Mellon University equipped Baxter with deep learning algorithms so that it could learn to grasp objects of various shapes and sizes [10]. Each of Baxter’s arms came with a standard two-fingered parallel gripper and a high-resolution camera that allowed the robot to see what was in its grasping range. Even without machine learning, Baxter’s repetitive motions could still lead to injury or property damage. But in this case, Baxter’s actions could be directly traced to human involvement in the system and thus determining liability would proceed under well-established principles of tort law [11]. However, with the ability to learn and to act with autonomy, if Baxter caused an injury or damaged property, under current law would a human still be liable for the robot’s actions? There are several proposed approaches to this problem which are discussed below.

To allow Baxter to learn how to grasp objects, Pinto and Gupta placed a variety of objects on the table in front of the robot and allowed it to operate for up to 10 hours a day without human supervision [10]. Baxter’s deep learning neural networks were pre-trained in object recognition, thereby giving Baxter some essential skills to start with. Additionally, two layers of the neural network were devoted to learning to grasp. After learning basic grasping skills, Baxter was exposed to multiple objects of various shapes and sizes, some of which were entirely novel. Over a period of a few weeks, Baxter performed some 50,000 grasps on 150 different objects, each time learning whether the approach was successful or not. With this learning Baxter was able to predict whether or not a grasp would be successful close to 80 percent of the time [10]. While a successful outcome, there is still plenty of room for improvement and a rather high probability that a grasp may be unsuccessful or even damage an object. If so, who is liable, the manufacturer of Baxter, the programmer, the person who sold Baxter to the end user, or person who wrote the algorithms controlling Baxter’s end effector? And if none of these parties were aware of the structure of the learning algorithms used by Baxter, should they still be considered liable for its actions?

In the backdrop of these questions involving autonomous robots, consider a neural network which is used to help robots learn new skills. When determining fault and liability for property damage or injuries to humans, I believe the algorithms controlling an autonomous robot, and particularly the hidden layers of the neural net where learning occurs are particularly relevant to the inquiry. Neural networks are arranged such that the actual operation, the weighting of probabilities, is not perceived by humans. In fact, as Curtis Karnow who has written extensively on law and intelligent machines points out, humans do not fix the way in which elements are weighted, and they usually do not even identify which elements are weighed [12]. The nets organize themselves and during training, the system makes adjustments to the nodes in the network, assigning more or less weight to input from earlier layers. While for some system architectures, the human may introduce gating functions, such as permitting/vetoing, machine learning essentially takes the human out of the decision-making loops controlling an autonomous robot, resulting in challenges to established tort law.

In addition to grasping objects, another essential task that humans perform with high accuracy is object avoidance as they navigate throughout the environment. As robots increase their mobility, they also increase the number of objects they encounter, this creates the possibility that the robot may damage property or cause injuries. The task of “navigating on the fly” has been difficult for robots to learn but progress is being made in this area. For example, Igor Mordatch, a postdoctoral fellow at the University of California, Berkeley, created a robot equipped with deep learning algorithms designed to aid the robot in navigating the environment in real-time ([13], see also [14]). The robot, Darwin, was programmed to learn by his mistakes and adjust his technique on the fly. Using algorithms that are modeled on the human brain Darwin uses two GPU-accelerated deep learning networks in which the learning takes place in hidden layers of the network. These neural networks strengthen or weaken connections between neurons in response to feedback.

Darwin’s learning took place in two stages: in simulation and in the real world. As a foundation for performance, Mordatch created a simulated model of Darwin’s physical presence (height, girth, etc.) and specified some basic properties of the environment. However, the robot was not directly taught to walk. In the simulation, the robot took what he knew and determined the right sequence of movements required for walking, such as how to position his legs to walk to a certain location or how to twist his torso to stand from a prone position. In the second stage of learning, Darwin applied what he learned in the simulation to stand, balance, and reach in the physical world. The mobility to explore the environment provided by deep learning algorithms will surely increase the possibility that Darwin will encounter novel objects and may be involved in an accident, thus raising the question of who is liable if the robot was acting autonomously and its actions were not able to be predicted beforehand by any human in the system?
3 Products liability, robots, and algorithms

When determining liability for an accident involving a robot that is sold as a product (i.e., not an “experimental robot” being developed in a laboratory), courts typically look to whether the robot was defective in some way. What law applies to products that are defective and lead to property damage or injuries? Under tort law, products liability is a legal theory which may be used to determine liability for property damage or injuries which involve products manufactured for sale; for example, industrial robots and home robots such as vacuum cleaners [11]. In common law jurisdictions, products liability claims can be based on a theory of negligence, strict liability, or breach of warranty depending on the jurisdiction within which the claim is based. In the U.S., the law of products liability is found mainly in state law and in the Uniform Commercial Code (UCC), and specifically Article 2 of the UCC which deals with the sales of goods. In the UCC two important products liability sections for our discussion are the implied and express warranties of merchantability in the sales of goods, that is, §§ 2-314, and 2-315 which states: “Where the seller at the time of contracting has reason to know any particular purpose for which the goods are required and that the buyer is relying on the seller’s skill or judgment to select or furnish suitable goods, there is unless excluded or modified under the next section an implied warranty that the goods shall be fit for such purpose.”

Many states in the U.S. have enacted comprehensive products liability statutes; however, in any jurisdiction one must prove that the product is defective—this may be difficult to do with autonomous robots that may operate successfully without a mechanical defect but still result in property damage or injuries due to the machine learning capabilities of the robot. There are three types of product defects that incur liability in manufacturers and suppliers of robots: design defects, manufacturing defects, and defects in marketing [11]. Design defects are inherent; they exist before the product is manufactured. While the item might serve its purpose well, it can be unreasonably dangerous to use due to a design flaw. On the other hand, manufacturing defects occur during the construction or production of the item. Only a few out of many products of the same type are flawed in this case. Finally, defects in marketing deal with improper instructions and failures to warn consumers of latent dangers in the product. On the last point, under product liability law, the developer of the algorithms controlling a robot can provide a warning in a number of ways, one would be an explicit and implicit documentation of the assumptions inherent in the algorithm, but this becomes difficult to implement as algorithms become more complex and unpredictable.

Under established law, the physical embodiment of a robot is considered property, it has no rights of its own, no matter how intelligent and independent it acts from human control. If a commercially developed robot is involved in an accident, the law which applies to a product placed in the stream of commerce is triggered. This is an important observation because many current autonomous robots are designed in research labs and not marketed as products, thus, they may not be considered a “product” under the UCC. This, of course, has implications on whether a cause of action can proceed under a theory of products liability. However, in my view, whether an algorithm or other analytical technique controlling an autonomous robot can be considered a product is an evolving topic under contract law. Clearly, robots that are manufactured and for sale are considered products and subject to product liability law, but what about the status of algorithms and other analytical techniques which may be provided to a robot manufacturer by a third party? If an algorithm is not considered a product, under what legal theory would compensation be possible for damages that result from decisions made by algorithms? It appears that the law as currently established may be useful for determining liability for mechanical defects, but not for errors resulting from the autonomous robot’s “thinking”; this is a major flaw in the current legal approach to autonomous robots.

Generally, products liability refers to the liability of any or all parties along the chain of manufacture and distribution of any product for damage caused by that product. This includes the manufacturer of component parts (at the top of the chain), an assembling manufacturer, the wholesaler, and the retailer (at the bottom of the chain). Inherent in product liability law is the notion that at some place within the distribution chain, there is a human who may have been a factor in a design or manufacturing defect or failed to provide a warning which eventually lead to an accident involving a robot. Additionally, with intelligent and autonomous robots controlled by algorithms, there may be no design or manufacturing flaw that served as a causative factor in an accident, instead the robot involved in an accident could have been properly designed, but based on the structure of the computing architecture, or the learning taking place in deep neural networks, an unexpected error or reasoning flaw could have occurred. In this case, who would be liable for any property damage?
or injuries sustained and would products liability be an applicable legal theory to analyze culpability of the actors?

As just discussed, for a product liability cause of action, the algorithm, or in some cases the software the algorithm is embedded in, must be a “product” (and not a service). Accordingly, two criteria are often presented. The first is that a product - as distinguished from a service - must consist of some physical embodiment that is available to the purchaser directly. Here one could argue that the algorithms controlling the robot are embedded within the software architecture or even embedded within the digital circuits controlling the robot. However, there is some debate as to whether software should be considered a product or a service. And second, the algorithm must not be a unique or specially designed item. For algorithms designed to control robots which are created in research labs, this second point may be a difficult bar to overcome. However, under case law, the concept of what is a product is evolving and the physical manifestation requirement, may not be dispositive. For example, water and electricity have been considered products by courts, so possibly algorithms and analytical techniques controlling a robot could be as well; although this remains to be determined. On this point, in the case, *Flour Corp. v. Jeppe- sen*, the court looked not to the dictionary meaning of a product but rather “within the meaning if it’s use” [15].

To some extent, as long as an algorithm or analytical technique used to control a robot is handcrafted in a university research laboratory, they are typically not marketed as products for sale, and thus they do not qualify as products under the UCC. This observation relates to the requirement that for a products liability cause of action, the defendant must be a seller of the product. Further, it is not unusual that algorithms and software programs controlling a robot are based on “toolkits” provided as open source to the robotics community by a research lab or even vendor. Individuals who program a robot based on routines provided by third party toolkits will not be able to bring a products liability cause of action for injury against the robot manufacturer because the algorithm that caused the injury was not the product (in this case, the physical robot) sold by the manufacturer.

### 3.1 Strict liability

In some cases, tort law imposes liability on defendants (such as manufacturers of a robot) who are neither negligent nor guilty of intentional wrongdoing. Known as strict liability, or liability without fault, this branch of torts seeks to regulate those activities that are useful and necessary but that create abnormally dangerous risks to society [11]. Strict products liability is predicated on the existence of an unreasonably dangerous product whose foreseeable use has caused injury. Whether an autonomous robot would be considered an unreasonably dangerous product remains to be determined as technology develops and more autonomous robots enter society; although military and police robots might in some circumstances meet this standard. Under some fact patterns, consumers who have been injured by defectively manufactured products may rely on a strict liability cause of action. And under the doctrine of strict products liability, a manufacturer must guarantee that its goods are suitable for their intended use when they are placed on the market for public consumption. The law of torts will hold manufacturers strictly liable for any injuries that result from placing unreasonably dangerous products into the stream of commerce, without regard to the amount of care exercised in preparing the product for sale and distribution and without regard to whether the consumer purchased the product from, or entered into a contractual relationship with, the manufacturer [11].

As another approach for liability, some commentators have suggested that the law which applies to domesticated animals could be adopted for intelligent robots [11, 16]. Generally, for animals considered to be “wild,” the courts apply strict liability for any damages that result from a “dangerous propensity” of the species. But among those who propose that animal law be applied to robots, robots with machine learning are far more comparable to domesticated than wild animals. For domesticated animals, strict liability only applies where the owner knew or had reason to know, that a particular animal (or type of robot for our discussion) had a tendency to bite or attack. For robots, the propensity to cause an injury could be ascertained from the algorithms used to direct the robot’s behavior; and if the owner or programmer of the robot had reason to know of the robot’s propensity for dangerous acts, liability for the robot’s actions would occur. I don’t particularly view the law of animals as being adequate for the actions of increasingly intelligent robots, because I believe their intelligence will eventually surpass that of domesticated animals and at some point, they may intentionally cause damage, but this section is included to provide another position on assigning liability for the actions of robots.

For a manufactured product, a plaintiff asserting a strict liability claim against a robot manufacturer must plead and prove, under a typical state’s law, that the defendant sold a product that was defective and unreasonably dangerous at the time it left the defendant’s hands; the product reached the plaintiff without substantial change; and the defect was the proximate cause of plaintiff’s in-
juries [11]. Additionally, strict liability cases do not impose liability on the manufacturer and other parts of the distribution chain for changes made to the product after delivery to the consumer unless those changes were foreseeable. Because autonomous robots, may “change” from their original programming, their actions may not be foreseeable. Foreseeability makes strict liability poorly suited for injury caused by autonomous robots.

The following case, and the cases in this section of the paper in general, are included to show examples of how the law has previously dealt with disputes involving robots. Even though the robots in these cases lack the autonomy of more recent robots equipped with machine learning, they provide a legal framework to consider future cases. In an interesting case, in Mracek v. Bryn Mawr Hospital [17] the plaintiff, Mracek brought an action against Bryn Mawr Hospital and Intuitive Surgical, Inc. for damages arising out of theories of strict product liability, strict malfunction liability, negligence, and breach of warranty in connection with a prostatectomy performed with an operative robot called the "da Vinci" which was manufactured by Intuitive. At issue was a motion for summary judgment filed by Intuitive (summary judgment is a judgment entered by a court for one party and against another party summarily, i.e., without a full trial). It is worth reviewing the case in more detail. Because of the robot’s malfunction, the surgical team abandoned its attempt at a robotic prostatectomy and switched to laparoscopic equipment to finish the procedure. At issue was whether the robotic surgeon was defective.

In the motion for summary judgment, Intuitive asserted that, since Mracek failed to submit any expert report that was critical of the da Vinci robot, he therefore could not meet his burden of proof with respect to any of the theories of liability that he had asserted, including strict liability. Mracek asserted that the defect in the da Vinci robot was obvious because it displayed an error message and had to be shut down during the operation. In disagreeing, the court reasoned that the da Vinci is a complex machine, one in which a juror would require the assistance of expert testimony (which Mracek failed to provide) in order to reasonably determine if the robot had a defect, without such evidence, the court granted summary judgment to the robotic manufacturer. Likewise, expert testimony will be necessary to convince a trier of fact that an autonomous robot had a defect that led to injuries. However, if the argument is that the defect was “cognitive,” rather than a mechanical or design defect, this may be difficult to prove just as it is with human defendants.

Interestingly, Mracek also brought a cause of action under a malfunction theory of products liability. Such a cause of action may be relevant where the harmed party is unable to produce direct evidence of a product’s defective condition, and thus, the precise nature of the product’s defect. This theory permits a plaintiff to prove a defect through circumstantial evidence. Under this theory, the plaintiff may raise a supportable inference of defect through: (1) evidence of the occurrence of a malfunction; (2) evidence eliminating abnormal use; and (3) evidence eliminating reasonable secondary causes for the accident [11, 17]. If such evidence is present, the plaintiff is relieved from demonstrating the precise defect in the product and the trier of fact is permitted to infer the existence of a defect from the malfunction and the negation of abnormal use and other reasonable secondary causes. But in the present case, Mracek offered no evidence to eliminate any reasonable secondary causes, and thus, he failed to meet his burden under the malfunction theory of strict product liability.

### 3.2 Negligence

Negligence is the term used in tort law to characterize behavior that creates unreasonable risk of harm to people and property [11]. Under the law, a negligence action is something a “reasonably prudent” person would not do. Thus, a person acts negligently when his behavior departs from the conduct ordinarily expected of a reasonably prudent person under the circumstances. An issue for autonomous robots is whether the “reasonably prudent person” standard would apply, or alternatively would there need to be a “reasonably prudent AI” standard adopted and if so, what would be the standard? (I believe it would be based on the algorithms used to control the robot). Further, a negligence claim may stand even in the absence of a "defect" under strict liability principles. A prima facie negligence claim requires the plaintiff to show that: (1) the defendant had a duty to conform to a certain standard of conduct; (2) the defendant breached that duty; (3) such breach caused the injury in question; and (4) the plaintiff incurred actual loss or damage [11].

Considering the actions of robots, under a negligent design theory, a plaintiff would need to show that a robotics manufacturer had a duty to exercise reasonable care in manufacturing the robot, the manufacturer failed to exercise reasonable care in making the robot, and the defendant’s conduct proximately caused plaintiff’s damages. Most injuries that result from tortious behavior are the product of negligence, not intentional wrongdoing. Interestingly, while we have a workable view of what it means for a person to act negligently or otherwise act
in a legally culpable manner, we have no similarly well-defined conception of what it means for an autonomous robot or algorithm to do so yet some robots are controlled by algorithms operating with no human supervision. Over time, the further that AI machines move away from pre-determined instructions and learn the more they will display behaviors that were not just unforeseen by their creators but were wholly unforeseeable. This is significant because foreseeability is a key ingredient for liability in negligence. Consider an advanced, fully autonomous, AI-driven robot that accidentally injures or kills a human, in this scenario, under current law, the lack of foreseeability could result in nobody at all being liable in negligence.  

In many ways, Jones v. W + M Automation, Inc., [18], is a typical product liability and negligence action involving a non-autonomous robot. In Jones, a robotic gantry loading system struck the plaintiff when he entered into an area behind a safety fence. The main issue litigated on summary judgment was whether the system was defective when the defendant sold it. General Motors had purchased the gantry loading system and then installed it without an interlock system which would have stopped the machine when people were present, thereby allowing employees to work on the system within the danger zone behind the safety fence while the system was operating. The plaintiff sued under theories of strict liability, negligence, failure to warn, and breach of warranty [18].

The court held, among others, that summary judgment was appropriate for the component manufacturer defendants under the “component part” doctrine, which states that a manufacturer of a non-defective component part of a product is not liable if its part is incorporated into another product that might be defective [18]. So, for robots, if the software, algorithms, or analytical capabilities are provided by a third party the robot manufacturer (if the physical robot had no defects) may not be liable for harm. The court also reasoned that the defendants were entitled to summary judgment because the plaintiff failed to introduce evidence in opposition to the summary judgment motion showing that the system was defective. In Jones, the court held that the danger involved with the plaintiff of going behind the safety fence while the system was in operation was an open and obvious one. Accordingly, the court held that summary judgment should have been granted to defendants on the failure to warn theory. One can image the difficulty of imposing a failure to warn requirement for robots that act with autonomy. Possibly, given the parameters of the algorithms controlling the robot, generic warnings could be developed. For example, people may need to be informed of the robot’s range of motions, its ability to navigate, its grasping ability and grasping strength, and so on.

Interestingly, several prior cases with robots have resulted in no liability being allocated against the robot manufacturer. Instead such cases focused on human culpability in accidents involving robots. More legal disputes involving robots can be expected to occur with more and more autonomous robots entering society; some in the role of companion, employee, or possibly as an independent contractor. As an example, in Payne v. ABB Flexible Automation, Inc., [19], a person was crushed and killed by a robot’s gripper arm in an auto wheel plant while working within the “cell” in which the robot operated. In Payne, the court found that the failure to meet American National Standards Institute (ANSI) standards, and the problems with unexpected movements by the robot admitted by the defendant manufacturer were irrelevant, because these problems did not cause the accident. Accordingly, the court affirmed summary judgment in favor of the manufacturer. The defendant, ABB, had submitted a report which indicated that inattention by Payne was the primary factor in the accident and found that Payne had overlooked safety measures by entering the cell before “locking it out,” and by running the robot at 100% test speed while inside the cell, rather than at 25% speed as required by safety guidelines. In summary, the plaintiff failed to produce evidence which showed that ABB’s failure to manufacture the robot with a safety device constituted negligence, or that it rendered the robot defective or unreasonably dangerous. The court reasoned that merely asserting that a safety device would have prevented an accident did not satisfy plaintiff’s burden of proving causation in the present case.

In a similar case, Miller v. Rubbermaid Inc. [20] the court affirmed the trial court’s entry of summary judgment in favor of the employer where: the record showed that the deceased employee placed himself in danger of being crushed by a robot used in plastic injection molding; the employer did not know of any dangerous condition of the machine; there was no evidence of safety violations; and the employer had not inadequately trained the decedent. Under these circumstances, there was no genuine issue of fact regarding intentional conduct by the employer. However, Behurst v. Crown Cork & Seal USA, Inc., [21], did not foreclose an intentional tort claim against an employer. The case arose out of an accident involving a blank transfer robot (moving metal from die to die) used in a can-making plant. The plaintiff was killed when, apparently, she was trapped in the danger zone when the machine restarted, and the access door closed behind her. The court found a jury question as to an intentional tort claim against an em-
ployer in light of: the employer’s knowledge of the flawed performance of the machinery and a history of prior accidents; the employer’s alleged refusal to reprogram the machine; the employer’s tolerance or encouragement of unsafe maintenance practices; the employer’s insistence on understaffing its production line; and the employer’s placement of the deceased alone in the production line without sufficient training. Making sense of the above cases in the background of increasingly intelligent and autonomous robots, clearly a factor in the analysis of determining liability will be the “human factor”; that is, the culpability of the human actor in placing themselves in harm’s way. However, as robots act more autonomously, it may be more difficult for humans to predict the robot’s actions, and thus to act with the appropriate precautions.

4 Algorithms, liability, and law

While the above cases focused on issues of liability for robots that operate mostly with predetermined rules and logic, the following discussion applies more so to autonomous robots and the challenges such robots offer established law when the issue is determining liability resulting from property damage or injury to humans. Note the point again that the above cases provide legal precedence for future cases involving autonomous robots. For machine learning robots that are controlled by algorithms, the issue of determining liability is complicated in several ways. For example, the algorithm’s performance may be based in part on the high-level language’s syntax and the semantics of a particular programming language. And every program below the level of the algorithm in the program hierarchy may have undetected programing or logic errors which could affect the performance of the robot. Given these layers of control, this makes the task of determining what actions initiated by the robot led to damages difficult.

Further, consider that all programs to control a robot’s behavior transfer inputs into outputs. When should the input supplied to the algorithm by the user be considered a substantial alteration which is a factor in determining liability under tort law, especially when the algorithm is designed to learn and alters its own “thought processes”? And when is the algorithmic reasoning process so changed that it has undergone a substantial alteration from the last time a human was involved in its design? If it can be shown that the algorithms or software controlling a robot can change its parameters and that change led to an accident, who then is liable? Further, when the buyer or user changes the way the logic used by the algorithm operates, extends the knowledge base used by the AI, or otherwise alters the basic nature of the program, the buyer or user would substantially alter the range of possible outcomes which has implications for liability.

Additionally, many aspects of “software law” are also relevant to discussions of liability for autonomous robots. In many cases software is owned through a license and typically contains boilerplate language freeing the software provider from virtually all forms of liability while binding the commercial user to various use restrictions. Legal theorists have looked outside of the contract realm for ways to hold software providers accountable for the harm that users sustain as a result of faulty code. For example, consumer protection laws would seem to offer one narrow avenue for redress. Alternatively, users have filed suit for compensation on tort grounds, alleging negligence on the part of the software provider or product defect. Specifically, in the U.S., commentators speculate that the contract-making principles embodied in the UCC - a set of model laws adopted at least partially by all the states - could be used to “pierce[e] the vendor’s boilerplate” and create a legal framework that would equally benefit vendors and users, licensors and licensees [22].

In fact, a commentator observed as early as 1985 that the courts have adequate means to protect software vendees from unconscionable contract provisions and the UCC makes requirements for effective disclaimers of warranty clear, so that the UCC could adequately protect software vendees and not serve as a vehicle for manufacturers to limit their liability. Unfortunately, the UCC has served more as a liability-limitation vehicle than offering consumers redress for faulty software. Treating software licenses as sales governed by the UCC creates a legal fiction which - contrary to the general intent of the UCC - places the purchaser at a severe disadvantage vis-à-vis the software vendor. For some time, the courts have accepted that operating assumption and allowed software providers to contract away responsibility for the deficiencies of their products. Will the same be the case for machine learning software as it becomes more available as a commercial product?

Several additional factors found in the standard license agreement prevent users from seeking compensation when they are harmed by defective software. To start, some software is free. This is a problem under contract law because courts will not hold software providers liable for harms brought about for products or services for which users did not offer some form of payment - or under contract law “consideration.” Further, software providers tend to prevail under a “private-ordering regime” and remain
immune even when users bring suit under various tort theories: the courts are resistant to finding an implied warranty of merchantability (this warranty guarantees that a product sold by a merchant will work when used for its intended purpose) with respect to software products and services that they know cannot be made vulnerability-free. Additionally, courts will likely continue construing software license agreements - and, as it turns out, tort actions - in favor of software providers.

With the above background in mind, consider the different types of software involved in the learning and control of autonomous robots. If the autonomous robot involved in damages is manufactured without mechanical flaws, to determine fault, courts will need to look carefully at the robot’s operating system, neural nets, deep learning algorithms, and other analytical techniques controlling the robot’s behavior. Clearly, it may be difficult to assign liability to any human if the party producing the injury was an autonomous robot guided by machine learning algorithms. Consider the following pseudocode which is provided only to emphasize the point that it is in the lines of the code itself that courts will likely look when determining whether an autonomous machine had the “mental state” to perform a negligent act or malpractice. And for neural nets, the weights and probabilities associated with learning will provide courts evidence of decision-making and the parameters used to classify data, such information could be analyzed by courts for various tort actions. For example, under negligence law, the court may examine the structure of an algorithm to determine whether it represents the actions of a “reasonably prudent person” in its decision making, or with modifications to the standard, a “reasonably prudent algorithm”; that is, an algorithm operating efficiently and appropriately for the task. Deviation from standards could result in liability.

```plaintext
startLocation = [7 22];
endLocation = [15 5];
% Search for a solution between start and end location.
path = findpath(prm, startLocation, endLocation);
while isempty(path)
    prm.NumNodes = prm.NumNodes + 10;
    update(prm);
    path = findpath(prm, startLocation, endLocation);
end
% Display path
Path
```

Of interest to this discussion is that computer scientists have developed ways to test the efficiency of algorithms (could such procedures be used to determine a “reasonable algorithm” standard?) and courts may take judicial notice of such procedures. To test an algorithm’s efficiency, one needs to define a test harness which consists, among others, of the data used to train the autonomous robot. In software testing, a test harness or automated test framework is a collection of software and test data configured to test a program unit by running it under varying conditions and monitoring its behavior and outputs. It has two main parts: the test execution engine and the test script repository. The goal of the test harness is to be able to quickly and consistently test algorithms against a fair representation of the problem being solved. The outcome of testing multiple algorithms against the harness is an estimation of how a variety of algorithms perform on the problem against a chosen performance measure and such tests could result in a standard being developed (which could be used in a tort action to define a reasonably prudent algorithm). The results will also give an indication of how learnable the problem is. If a variety of different learning algorithms universally perform poorly on the problem, it may be an indication of a lack of structure available for algorithms to learn. This may be because there actually is a lack of learnable structure in the selected data or it may be an opportunity to try different transforms to expose the structure to the learning algorithms.

The performance measure is a way to evaluate a solution to the problem and thus could have relevance for discussions of liability for autonomous robots. It is the measurement that is made of the predictions made by a trained model on the test dataset. In the context of discussions on liability, performance measures will likely be specialized to the class of problems an autonomous robot will perform. Additionally, many standard performance measures will give a score that is meaningful to the problem domain. For example, classification accuracy may be calculated as the total correct corrections divided by the total predictions made. For determining liability, it may also be useful to gain a more detailed breakdown of performance, for example, to learn about false positives if they occur. There are many standard performance measures to choose from and the courts will have to sift through them.

## 5 Summary

With a lack of legal rights provided for autonomous robots, under current law those who write algorithms and develop analytical techniques that control a robot’s behavior, are parties to be considered in the chain of liability leading to damage or injuries. This might involve applying a the-
ory of strict liability – that is, liability applied without any
need to prove fault or negligence for any harms resulting
from the actions of autonomous robots. As noted in this
paper, strict liability already exists for defective product
claims in many jurisdictions and for the actions of ani-
mals, therefore extending it to the actions of autonomous
robots would be one way to allocate liability. However,
the issue of foreseeability of the type of harm experienced
(that is, were the damages foreseeable), will be a defense
raised by the defendant in any strict liability tort action.
Alternatively, as some have suggested, there could be a no-
fault liability scheme with a claims pool contributed to it
by the AI and robotics industry that would be available to
pay damages when an autonomous robot was involved in
an accident and no human was identified as a factor in the
accident (see generally [23]).

On the regulatory side, development of rigorous safety
standards and establishing safety certification processes
for algorithms will be necessary as more autonomous
robots enter society [24]. To create a suitable framework
of institutions and processes to focus on damage and injuries
resulting from autonomous robots, input from robotics
and AI experts will be needed to help establish any regula-
tory framework. This is due to the complexity of machine
learning techniques and the general lack of understand-
ing of machine learning techniques outside the AI and
robotics R&D community. This also means that advisory
committees to legislatures and governments should be es-
tablished to help determine how to regulate autonomous
robots. And any framework developed will also have to be
flexible enough to take account of both local jurisdictional
considerations and global considerations; possibly con-
sisting of mutual recognition of safety standards and cer-
tification between countries, and the need to comply with
any future international treaties or conventions.

When regulating the actions of robots, legislators will
need to consider that some may still have human involve-
ment with the robot at least at a supervisory level. Inter-
estingly, if robots continue to gain in performance and ac-
curacy, the failure to use a robot may someday actually
create liability. But even with autonomous robots, humans
may still have a duty of care to avoid accidents, even after
delegating some of the operation to an autonomous robot.
This liability has already been a factor in the use of semi-
autonomous machines. For example, in an early case, in
Brouse v. United States, [25], the court held that a pilot had
a duty to be on the lookout to prevent air-to-air collisions
while flying under “robot control” (or autopilot).

For robots operating predominantly without auton-
omy, a robot manufacturer may be held liable for negligent
failure to warn if the manufacturer knows or has reason
to know that: 1) the product is likely to be dangerous for
its intended use, 2) it has no reason to believe that users
will realize its dangerous condition, and 3) it fails to ex-
cise reasonable care to inform users of its dangerous con-
dition. Nonetheless, if the danger is open and obvious, the
manufacturer may have no duty to warn users. Addition-
ally, the actions of autonomous robots devising their own
means to attain goals may not be subject to liability un-
der any theories of tort, to the extent they are adoptive
because this ability may not allow foreseeable harm to be
predicted. And if the algorithms and analytical techniques
that are used to control autonomous robots are not con-
sidered products, perhaps service liability may be a cause
of action evolving from products liability. This may be ap-
licable where the line between a service and product is
unclear which seems to be the case with algorithms em-
bedded within robots performing a range of “service-like”
activities.

Just as the actions of human professionals may result
in malpractice, some have argued that a similar cause of
action be developed for computers and more recently for
AI entities and intelligent robots. However, thus far, an
attempt to create a “computer malpractice” cause of ac-
tion has failed; so too might such a cause of action fail for
robots. Malpractice which is currently unavailable for al-
gorithms that learn, imposes liability for negligence and
is judged by a higher standard than that of a reasonable
person, it judges according to the standard of a profes-
sion’s average practitioner [11]. The base standard is set by
the professional community. Should there be a malprac-
tice cause of action for algorithms? Traditionally this cause
of action has been restricted to professions which require
judgment and independent analysis and that is what ma-
cine learning algorithms are designed to do. Similarly, is
the development of algorithms (e.g., by humans or neural
nets) a profession to be judged under malpractice law for
liability? If so, the AI expert could be liable for malprac-
tice. But, of course under current law algorithms are not
subject to malpractice claims and are not considered to be
professionals.

Another relevant comment for liability in discussions
of autonomous robots, is provided by judge Curtis Karnow
who proposes the concept of an “electronic personal-
ity” [26] (see the European Union’s discussion of the con-
cept [27]). He notes that corporations, partnerships, and
associations have substantive rights, and also have proce-
dural rights to bring suit. Similarly, if autonomous robots
were granted rights, perhaps they could be held liable for
their actions, opening up several tort theories to evaluate
their liability. Rights for robots may happen eventually,
but for now, determining liability for autonomous robots
remains difficult and under some circumstances impossible, thus motivating the need for more debate and legislative action.

References