Volume 4, No. 5 | September–October 2021

RAIL

The Journal of Robotics, Artificial Intelligence & Law

fastcase FULL COURT PRESS

RAIL
The Journal of Robotics, Artificial Intelligence & Law
Volume 4, No. 5 | September–October 2021

THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW (ISSN 2575-5633 (print) /ISSN 2575-5617 (online) at $495.00 annually is published six times per year by Full Court Press, a Fastcase, Inc., imprint. Copyright 2021 Fastcase, Inc. No part of this journal may be reproduced in any form—by microfilm, xerography, or otherwise—or incorporated into any information retrieval system without the written permission of the copyright owner. For customer support, please contact Fastcase, Inc., 711 D St. NW, Suite 200, Washington, D.C. 20004, 202.999.4777 (phone), 202.521.3462 (fax), or email customer service at support@fastcase.com.

Publishing Staff
Publisher: Morgan Morrissette Wright
Journal Designer: Sharon D. Ray
Cover Art Design: Juan Bustamante

Cite this publication as:

The Journal of Robotics, Artificial Intelligence & Law (Fastcase)

A Full Court Press, Fastcase, Inc., Publication

Editorial Office

711 D St. NW, Suite 200, Washington, D.C. 20004
https://www.fastcase.com/

POSTMASTER: Send address changes to THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW, 711 D St. NW, Suite 200, Washington, D.C. 20004.

Articles and Submissions

Direct editorial inquiries and send material for publication to:

Steven A. Meyerowitz, Editor-in-Chief, Meyerowitz Communications Inc., 26910 Grand Central Parkway, #18R, Floral Park, NY 11005, smeyerowitz@meyerowitzcommunications.com, 646.539.8300.

Material for publication is welcomed—articles, decisions, or other items of interest to attorneys and law firms, in-house counsel, corporate compliance officers, government agencies and their counsel, senior business executives, scientists, engineers, and anyone interested in the law governing artificial intelligence and robotics. This publication is designed to be accurate and authoritative, but neither the publisher nor the authors are rendering legal, accounting, or other professional services in this publication. If legal or other expert advice is desired, retain the services of an appropriate professional. The articles and columns reflect only the present considerations and views of the authors and do not necessarily reflect those of the firms or organizations with which they are affiliated, any of the former or present clients of the authors or their firms or organizations, or the editors or publisher.

QUESTIONS ABOUT THIS PUBLICATION?

For questions about the Editorial Content appearing in these volumes or reprint permission, please contact:

Morgan Morrissette Wright, Publisher, Full Court Press at mwright@fastcase.com or at 202.999.4878

For questions or Sales and Customer Service:

Customer Service
Available 8 a.m.–8 p.m. Eastern Time
866.773.2782 (phone)
support@fastcase.com (email)

Sales
202.999.4777 (phone)
sales@fastcase.com (email)
ISSN 2575-5633 (print)
ISSN 2575-5617 (online)

Defining Autonomy in the Context of Tort Liability: Is Machine Learning Indicative of Robotic Responsibility? Part I

Katherine D. Sheriff*

Traditional tort law benefits consumers by holding accountable parties responsible for injury, encouraging greater care in manufacture and making injured victims whole. Whether traditional notions of legal responsibility comport with the advent of artificial intelligence has been the sweetheart of academic research for decades. But less attention has been given to the comparatively limited role learning patterns and brain functionality play in the changing landscape of robotic service products. This article outlines traditional notions of learning and brain activity juxtaposed against software-engineered machine learning. It then describes the existing scholarly debate surrounding the issue of robotic autonomy and explains how this debate impacts the applicability of tort liability to robots.

Framing the Current Debate

Traditional thinking in tort, contract, and criminal law focuses on the responsibilities attributed to parties by considering intent.[1]

In tort law, the level of intent is more abstract and can be reduced to reasonableness, or awareness, in most cases.[2]

In contract law, the primary concern with contract enforceability is whether valid consideration was provided—whether there was a "meeting of the minds," necessitating awareness of the bargained-for terms.[3]

In criminal law, intent is most concretely applied as the *mens rea* affecting gradation of criminal culpability, apart from strict liability crimes, a binary system.[4]

The advent of "machine learning" is effectively spreading the reach of liability beyond the traditional notions applied to this point.[5]

Deep learning is not treated as fully in this article as might be necessary to provide a full evaluation in terms of autonomous agents. Complexity has been defined as the number of computations necessary relative to the problem size or the thing accomplished

Robotics, Artificial Intelligence & Law / September–October 2021, Vol. 4, No. 5, pp. 353–372.

ISSN 2575-5633 (print) /ISSN 2575-5617 (online)

and explaining that human creativity is the interconnection of dispersed or disparate thoughts.

No one really knows what a computer is going to do. There is no line of code or measure of the ability to create. Rather what the computer creates shows what it understands. For example, the Lovelace 2.0 Test of Artificial Creativity and Intelligence measures the ability to understand, not the ability to create.[6]

Liability in a tort sense does not automatically follow this expansion in scope to include robotic awareness in light of newly available technology. One reason for this disconnect is that discussion of the social benefits resulting from artificial intelligence ("AI") does not address the social loss from potential manufacturer immunity. Lawmakers have yet to account for the changing calculus introduced by a new brand of player in the liability "game." The more machine learning is accepted the more it will exacerbate the disconnect from traditional tort liability in which decisions are assessed according to a fault scheme that does not account for the possibility that proving black box causality will be prohibitively expensive.[7]

Using current neuroimaging and organizational psychology, this article seeks to show that decision making does not occur as traditionally portrayed and, given this reality, the current tort scheme must expand definitions of accountability.[8]

To begin, defining a legal term is an abstract task. When considering questions in an abstract sense, asking four distinctive questions proves helpful in laying out the framework for discussion. What must be defined? Who are the players? What functions do the players play? What functions should the players play?

Identifying the players in this particular problem is not difficult. Broadly, the players are the innovators, academics, lawyers, and consumers participating in the service products market. How the players currently function and in what roles is also fairly simple to explain.[9] The normative question of what should happen to remedy the problem identified is closely connected with the definitional question establishing what it is that we wish to use as the new liability standard.[10]

Defining concepts in this vein is an arduous task. Although not a full list, such concepts calling for more careful definition include liability, intent, deliberation, AI, machine learning, legal personhood, awareness, decision making, competence, and foreseeability.

The first part of this article outlines traditional notions of learning and brain activity juxtaposed against software-engineered machine learning. It then describes the existing scholarly debate surrounding the issue of robotic autonomy and explains how this debate impacts the applicability of tort liability to robots.

In the second part, to be published in an upcoming issue of *The Journal of Robotics, Artificial Intelligence & Law*, this article argues that because situations exist that cannot be pre-programmed, there is a strong case for mandated insurance carried by all robots to protect injured victims in those situations where fully autonomous machines "choose" a course of action too far removed from the work of the original programmer.

Then, this article recommends a variation of Ugo Pagallo's "digital peculium" liability scheme to account for the situations just described, where fully autonomous robots make decisions absent the appropriate linkage to the original creator and outside the scope of the uncertainty already programmed.

Finally, this article concludes by looking first to the allocation of systemic risk as a matter of organizational psychology and second the location of the hard case discussed in this article in the larger abstraction laid out by Ugo Pagallo and HLA Hart and Ronald Dworkin before him—the determination of a right answer or conclusive indetermination of any answer to robotic application queries.

Machine Learning: Robotic Potential to "Learn" and Decide

To begin, "AI is basically smart software that enables machines to mimic human behavior."[11] For example, using computer algorithms to "learn" over time, email uses machine learning to figure out from watching user behavior which emails are spam and which should be sent to the user's email inbox.[12] The concept of an intelligent agent is used to provide concreteness to the abstract world of AI. Agents are to receive percepts from the environment and then perform actions according to such input. By mapping percepts to actions in various ways, agents function as "production systems, reactive agents, logical planners, neural networks and decision-theoretic systems."[13]

While both machine learning and AI work together in consumer applications, the former centers around the creation of data-generalizing algorithms directed to improve performance and behaviors.[14] AI encompasses a broad scope beyond machine learning into areas like natural language processing and understanding.[15]

Machine-to-machine ("M2M") communication describes technologies allowing wireless systems to communicate with other devices of the same type. Other terms of some relevance are the "Internet of Things," the "Industrial Internet," and "Big Data."

Machine learning works to extend the scope of the designer's agent into unknown environments. In this way, it is possible to illustrate through machine learning how such capability constrains design to favor explicit knowledge representation and logical reasoning.

Robotics then is a field tied to vision for service. Two questions that push to the forefront in this field: How can one construct computer systems that automatically improve through experience and what are the fundamental statistical-computational-information-theoretic laws that govern all learning systems, including computers, humans, and organizations?[16]

Decision theory is the normative theory that asks how a rational agent should act.[17] Decision theory is the sum of probability theory and utility theory.[18] Applying economic theory would be enhanced if it were also descriptive; that is, asking how actual human decision making operates.[19] The fundamental idea supporting decision theory is the principle of Maximum Expected Utility ("MEU"): "An agent is rational if and only if it chooses the action that yields the highest expected utility, averaged over all the possible outcomes of the action."[20]

Artificial Intelligence: Defining Intelligent Agents

Because there are various ways to think of AI, it is important to designate at the outset whether AI will be treated as concerns with thinking or behavior and whether the goal is to model humans or to work from an ideal standard.[21] According to Stuart J. Russell and Peter Norvig, an agent is "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through effectors."[22] AI is generally organized into four categories of varying definitions.[23]

Systems that think like humans.	Systems that think rationally.
Systems that act like humans.	Systems that act rationally.

As Russell and Norvig explain, "The crucial part of executing conditional plans is that the agent must, at the time of execution of a conditional step, be able to decide the truth or falsehood of the condition—that is, *the condition must be known to the agent* at that point in the plan."[24]

But first, "the conditional planning agent can sometimes do better than the standard planning agents described earlier [in the chapter].

"Furthermore, there are cases where a conditional plan is the only possible plan. If the agent does not know if its spare tire is flat or inflated, then the standard planner will fail, whereas the conditional planner can insert a second conditional step that inflates the spare if necessary.

"Lastly, if there is a possibility that both tires have holes, then neither planner can come up with a guaranteed plan. In this case, a conditional planner can plan for all the cases where success is possible and insert a *Fail* action on those branches where no completion is possible."[25]

Approaching knowledge engineering for probabilistic reasoning systems is similar to how one might approach logical reasoning systems.[26]

A belief network is represented graphically below.[27]

1. A set of random variables makes up the nodes of the network.
2. A set of directed links or arrows connects pairs of nodes. The intuitive meaning of an arrow from node *X* to node ***Y*** is that *X* has a direct influence on *Y*.
3. Each node has a conditional probability table that quantifies the effects that the parents have on the node. The parents of a node are all those nodes that have arrows pointing to it.
4. The graph has no directed cycles (hence is a directed, acyclic graph, or DAG).

Dynamic decision networks are attractive because such systems provide potential solutions to the problems identified in AI systems that originate in controlled, test environments unfit for the complex uncertainty of the real world.

A schematic design for rational agents is illustrated below as "a decision-theoretic agent."[28]

function Decision Theoretic Agent (*percept*) **returns** *action*
calculate updated probabilities for current state based on
   available evidence including percept and previous action
calculate outcome probabilities for actions
   given action descriptions and probabilities of current states
select action with highest expected utility
   given probabilities of outcomes and utility information
return *action*

In practice, actions are chosen by the robot, or algorithmic software within the robot, calculating the parameters of the particular "decision network" for each potential setting of the appropriate "decision node." Nodes represent points where the decision maker has choice among two or more actions.[29]

After setting the decision node, the decision node then acts as a chance node set as an evidence variable.[30] The following is an algorithm for evaluating decision networks.

1. Set the evidence variables for the current state.
2. For each possible value of the decision node:
   a. Set the decision node to that value.
   b. Calculate the posterior probabilities for the parent nodes of the utility node, using a standard probabilistic inference algorithm.
   c. Calculate the resulting utility for the action.
3. Return the action with the highest utility.

Russell and Norvig explain that the decision network algorithm "is a straightforward extension of the belief network algorithm and can be incorporated directly" into the agent design.[31]

Machine Leaning: Algorithmic Knowledge Engineering

One of the most rapidly growing technical fields, machine learning explores how computers improve automatically through experience.[32] The general model of learning agents incorporates components of the performance element, or representations of the components, including available feedback and prior knowledge.[33]

Learning the performance element of an agent can be achieved in multiple ways, including "a direct mapping from conditions on the current state to actions; a means to infer relevant properties of the world from the percept sequence; information about the way the world evolves; information about the results of possible actions the agent can take; utility information indicating the desirability of world states; action-value information indicating the desirability of particular actions in particular states; and goals that describe classes of states whose achievement maximizes the agent's utility."[34]

Modeled using a "Skeleton for a simple reflex learning agent" Russell and Norvig explain the "learning element just stores each example percept/action pair. The performance element either does whatever was done last time for a given percept or induces an action from similar percepts.

The set of examples is a global variable that is shared by the learning and performance elements."[35] The following is a "Skeleton for a simple reflex learning agent" as modeled by Russell and Norvig.

global *examples* - {}
function Reflex-Performance-Element (*percept*) **returns** an action *if* (*percept*, *a*) in *examples* **then return** *a* **else**   *h* - Induce (*examples*) **return** *h* (*percept*)
procedure Reflex-Learning-Element (*percept*, *action*) **inputs**: *percept*, feedback percept   *action*, feedback action *examples* - *examples* U {(*percept*, *action*)}

Russell and Norvig describe learning decision trees as performance elements within the broader discussion regarding the expressiveness of decision trees.[36] The decision tree learning algorithm is modeled below.[37]

function Decision-Tree-Learning (examples, attributes, default) **returns** a decision tree
inputs: *examples*, set of examples
  attributes, set of attributes
  default, default value for the goal predicate
if *examples* is empty **then return** *default*
else if all *examples* have the same classification **then return** the classification
else **if** *attributes* is empty **then return** Majority-Value (*examples*)
else
  best - Choose Attribute (*attributes*, *examples*)
  tree - a new decision tree with root test *best*
  for each value v_i of best **do**
  $examples_i$ - {elements of *examples* with *best* = v_i}
  subtree - Decision-Tree-Learning (*examples*, *attributes* - *best*,
  Majority-Value (*examples*))
  add a branch to *tree* with label v_i and subtree *subtree*
end
return *tree*

In computational learning theory, the primary issue is one of quantity and sufficiency in terms of how many examples are needed.[38] Russell and Norvig explain, "Learning means behaving better as a result of experience."[39] Computational learning theory is based on the idea that any hypothesis that is "seriously wrong" will be "found out" with high probability after a small number of examples when it makes an incorrect prediction.[40] Russell and Norvig model the concept below in "an algorithm for learning decision lists."[41]

function Decision-List-Learning (*examples*) **returns** a decision list, *No* or failure
if *examples* is empty **then return** the value *No*
t - a test that matches a nonempty subset $examples_t$ of *examples*
  such that the members of $examples_t$ are all positive or negative
if there is no such *t* **then return** failure
if the examples in $examples_t$ are positive **then** *o* - Yes
else *o* - No
return a decision list with initial test *t* and outcome *o*
  and remaining elements given by Decision-List-Learning (*examples* - $examples_t$)

The primary issue then is how large a sample set needs to be in order to determine the algorithm is making an incorrect prediction.

Neural Networks: Brains versus Digital Computers in the Robotics Industry

Comparing the human brain with digital computers as a functional matter is instructive in determining the scope of robotic decision making. Industry welcomed AI and perpetuated development by prioritizing machine learning in existing products.[42] Google CEO Sundar Pichai explained, "We're thoughtfully applying [machine learning] across all our products, be it search, ads, YouTube, or Play. We're in the early days, but you'll see us in a systematic way think about how we can apply machine learning to all these areas."[43] At the time, robotic capabilities included seeing, reading, listening, talking, and writing.[44]

Human decision-making terminology includes the following: "Each of the processes could be dissociated in the fMRI analysis."[45] The "expected reward" is how much could be obtained for a given physical force, correlated with increase activation in the caudate as well as CMA (cingulate motor area) and SMA (supplementary motor area).[46] The "effort evaluation" is associated with any change in activation in the basal ganglia (nucleus accumbens, caudate, putamen), as well as the CMA.[47]

So, an increase in the effort requirement produced a decrease in activation.[48] "Stake evaluation" is activated in the right ventrolateral frontal regions often implicated in directing attention.[49] Finally, "cost-benefit weighing"—which becomes more difficult as expected costs come close to possible benefits—is positively correlated with activation in a network of brain regions previously associated with cognitive control on the medial frontal wall, including dACC (dorsal anterior cingulate cortex) and pre-SMA.[50] "Cost-benefit weighing" is negatively correlated with activation in ventromedial prefrontal cortex.[51]

Autonomous learning achieved through reward-modulation was developed in a recent study of human learning capabilities, in which a variation of the exploratory Hebbian ("EH") rule, which test neuron activity in varying scenarios in light of pre- and post-synaptic neural activity, and neuromodulators like dopamine.[52] The study varied from the traditional model by removing the third factor and the information-processing requirements to which the third factor applies.[53]

The Liability Framework: Accountability for Autonomy

It is useful to note the current scholarly debate and those reigning arguments existing in the debate we seek to join. To that end, the academics include Turing, Searle, Solum, and Berkeley in the following table of scholarly predictions of robotic cognitive abilities.[54]

Table 1: Scholarly Predictions of Robotic Cognitive Abilities

Academic/ Concept	Awareness/ Perception	Understanding/Predictability	Intent/ Deliberation	Implications
Turing Imitation Game	Uncertain	Yes. In discrete-state machines with input signals.	Sort of. Machines intended to carry out human operations.	Beginnings

Academic/ Concept	Awareness/ Perception	Understanding/Predictability	Intent/ Deliberation	Implications
Searle Chinese Room	No. Free will is the "something" that AI are missing and would make them indelibly human.	No. Robots do not understand the actions performed.	No. Robots could not "intend" what they could not understand. Robots need some biological structure.	Children; mentally impaired individuals; legal competency; guardianship
Solum 13th Amendment	Yes. If free will occurs through conscious reasoning and deliberation, then AI could possess free will.	Yes.	Yes.	Add whatever is necessary
Berkeley Many Mansions	Yes, eventually.	Yes, eventually.	Yes, eventually.	Mental processing; mental learning

The Turing Test is a method of inquiry by Alan Turing that determines whether a computer has the ability to exhibit intelligent behavior indistinguishable from human behavior. The Chinese Room Argument is a well-known thought experiment by John Searle. Of the many responses to the Chinese Room Arguments, two selected for discussion are Lawrence Solum's 13th Amendment analogy and the Berkeley Many Mansions Reply.

In the Chinese Room Argument, Searle envisions circumstances in which he is hidden in a room and presented with an "input" card questions, which are written in Chinese:

> Searle knows no Chinese; indeed, he is quite unaware of the enterprise in which he is engaged and is ignorant of the fact that the strange marks on the cards represent questions framed in Chinese. He consults a manual telling him (in English) precisely what equally strange marks to write on an "output" card, which he posts back to the outside world. By virtue of the "machine intelligence" embodied in the manual (which is

> actually a formalisation of the steps in an AI program), these marks on the output card constitute an answer to any input question. To a Chinese speaker external to the room, by virtue of its question answering ability, the system passes the Turing test for machine intelligence, yet the system implemented by Searle-in-the-room is entirely without understanding simply because Searle understands nothing.[55]

One interesting take on Searle's Chinese Room is the argument that all humans have free will and the exercise of such free will is the "something" that AI are missing that would make them indelibly human. First, is an AI human or if not is it because AI is missing some key human element? Is AI better conceived of as property?

Solum distinguishes between dependent and independent legal personhood. Specifically, AI is missing some key human element inherent in all warranting protection under the 13th Amendment of the U.S. Constitution.[56] What Solum says about Searle's Chinese Room might make sense when considered from the perspective that what robots are missing is deliberation—the ability to intend an action independent of the direction of another.

Assuming Searle is correct that robots do not "understand" the actions performed, then Searle could also be correct that the same robots could not "intend" what they could not understand. The implications for this argument for children, mentally impacted individuals, legal competency, and guardianship stretch beyond the immediate horizon. Perhaps this is why current scholars are not tackling the ideas at full speed or such ideas have been dismissed based on machine learning developments.

The many mansions reply was identified by Searle in his 1980 article. The concept is that there are many possible kinds of computer and computation. In the future there could be computers that would display AI.[57]

It might be easy to agree with the Berkeley "many mansions" reply because humans have developed machines with causal processes. However, the hard question remains of what is left when the argument is no longer directed at thinking ability and the explanatory powers of exposition. Ultimately, Solum lets the air out of the free will argument bubble by steering the idea toward definitions of autonomy.[58]

To switch gears focusing exclusively on autonomous vehicles, the objections have little merit base on the technology involved in driverless systems. Autonomous vehicles perceive, understand, and "learn" via integration of gathered information into existing data. Autonomous vehicles also might make "value judgments" in a sense. The free will argument is an interesting starting point as it typically involves arguments with religious undertones. The possibility of incorporating the argument from the view of souls versus soulless beings is esoteric but fascinating. Alternatively, using current neuroimaging or organizational psychology to demonstrate that decision making as popularly portrayed.

Legal Accountability as a Function of Robotic Intentionality

According to Searle, intentionality is qualitatively different in human beings, animals, and Martians because it is "a product of causal features of the brain" Searle assumes "is an empirical fact about the actual causal relations between mental processes and brains."[59] Searle's main premise is that certain brain functions are sufficient to establish intentionality. Searle's main argument is rooted in the idea that instantiating a computer program is never sufficient to establish intentionality:

> It is not because I am the instantiation of a computer program that I am able to understand English and have other forms of intentionality . . . but as far as we know it is because I am a certain sort of organism with a certain biological (i.e., chemical and physical) structure, and this structure, under certain conditions, is capable of producing perception, action, understanding, learning, and other intentional phenomena. . . . Martians also have intentionality but their brains are made of different stuff.[60]

To an external observer, an AI program without the ability to understand could still give the impression of intelligence.

Lawrence Solum provides a different perspective:

> The fact that human neural systems operate on the basis of a combination of electrical transmissions and biochemical processes does not make them any less subject to the laws of physics than are computers. The most plausible story about human free will is that an action is free if it is caused in the

> right way—through conscious reasoning and deliberation. But in this sense, AIs could possess free will.[61]

Further discussion is more readily facilitated within the context of contract law.

Contract Law

Hard cases within the realm of contract law revolve around robotic accountability relative to traditional contract law. Interestingly, contract law is based largely on expectations. In this context, Tom Allen and Robin Widdison distill down an argument by Günter Teubner that intelligences capable of autonomy are those with the social capacity to warrant granting of legal contracting power or recognition: "Hence, we can translate Teubner's point by stating that it makes legal policy sense to grant legal capacity to information systems that already have social capacity for autonomous action."[62]

Daniel Dennett describes intentional systems theory as a matter of semantics, "In traditional parlance, we seem to be attributing minds to the things we thus interpret."[63]

In contract law as put forth by the standard used in Allen and Widdison's thought experiment a "meeting of the minds" is necessary for contract formation.[64] Allen and Widdison relate the "standard, classical statement of the requirements for contract formation" as requiring:

- "[T]wo or more separate and definite parties to the contract";
- Said "parties must be in agreement";
- Said parties must "intend to create legal relations in the sense that the promises of each side are to be enforceable" as a function of promises; and
- Such "promises of each party must be supported by consideration."[65]

Benefits of attributing contracting power to robots include allocation of risk in an arguably more efficient manner than is typically the case with human actors. The absurd result where a robot acts erratically in conflict with existing societal norms and business practices could be rectified by holding human actors accountable for such robotic actions.

Arguably assigning responsibility to human actors only makes sense where it would be foreseeable that an agreement could not be based on a flagrant mistake. Dennett's intentionality scheme seems most relevant in cases where human actors and robots could escape contractual liability because robots could not possess the knowledge to truly intend to enter into a contract or consciously accept an offer.

Legal Accountability as a Function of Robotic Consciousness

Capable of being traced back to German automation scholars in the late 1800s, the contemporary debate on cognitive automata as software agents represents the traditional argument that "robots are tools, incapable of serving as agents themselves."[66]

According to Tom Dietterich while president of the Association for the Advancement of Artificial Intelligence, AI presents a threat when AI systems are completely autonomous:

> A misconception I think people have is that somehow these systems will develop free will on their own and become autonomous even though we didn't design then with that in mind. . . . AI systems will not become spontaneously autonomous; they will need to be designed that way. So, I think the danger of AI is not so much in artificial intelligence itself, in its ability to reason and learn, but in the autonomy.[67]

To Dietterich, implementation of autonomous vehicles and "Da Vinci robots" must be preceded by fully addressing risks associated with use, "By definition a fully autonomous system is one that we have no control over. And I don't think we ever want to be in that situation."[68]

Susan Schneider described the transition, arguing that the neuroscience discovery of algorithms in the brain underlying computation will convince scientists that brains are actually computational entities.[69]

To Schneider it follows that future persons could live with hybrid minds—part "natural" and part artificial. In this way, Schneider proposes that scientists reverse engineer the human brain to produce AI running with human brain algorithms. After AI operate with human brain algorithms, "Other AI creatures could have minds that are entirely different, borrowing from sensory modalities that other animals have (e.g., echolocation), featuring radically enhanced working memory capacity, and so on. Existing human brains could be enhanced in these novel ways as well. In sum, a plurality of distinct sorts of artificial minds could be sculpted."[70]

Robert M. French notes the limits of the Turing Test as physical, "One of the tacit assumptions on which Turing's proposed test rests is that it is possible to isolate the 'mere' (and thus unimportant to the essence of cognition) physical level from the (essential) cognitive level. This is the reason, for example, that the candidates communicate with the interrogator by teletype, that the interrogator is not permitted to see them, and so on."[71] To the point, French explained the inherent link between the physical and cognitive:

> Subcognitive questions, however, will always allow the interrogator to "peek behind the screen." The Turing Test is really probing the associative concept (and sub-concept) networks of the two candidates. These networks are the product of a lifetime of interaction with the world which necessarily involves human sense organs, their location on the body, their sensitivity to various stimuli, etc. Thus, while no one would claim that the physical location of eyes had anything essential to do with intelligence, a Turing Test could certainly distinguish this individual from a normal human being. The moral of the story is that the physical level is not dissociable from the cognitive level.[72]

As a practical matter, considering the objections within the context of tort law elucidates the impact on agency theory.

Tort Law

Increasing autonomy in what is becoming household AI and Internet of Things presses the boundaries of agency theory in which robots have been only tools through which human actors operate and are ultimately liable. Some elements of tort law come into play when rationalizing liability for robots to a certain extent, specifically the cheapest cost-avoider principle.

The complexity introduced by robotic applications in contractual promises actually simplifies the debate by making moot the most common rebuttals asking whether a robot can act lawfully, and if so whether a robot acting unlawfully comes back to the manufacturer or robot.

Legal Accountability as a Function of Robotic Predictability

Accountability based on autonomy is discussed indirectly by Curtis Karnow.[73] Describing results of "the evolutionary development of a real, neural-network driven mobile robot," Karnow details the evolutionary approach to understanding neural controllers for autonomous agents. Karnow notes that this approach has been successfully used by many researchers, but most studies use computer simulations. Against this backdrop, Karnow evaluates the evolutionary process taking place entirely within a robot without human contribution.

Karnow's discussion of experiments identifies "a simple task of navigation and obstacle avoidance" involving "a number of emergent phenomena that are characteristic of autonomous agents," explaining:

> We have neither pre-designed the behaviors of the robot, nor have we intervened during evolution. The robot itself and alone has developed—starting from a sort of tabula rasa—a set of strategies and behaviors as a result of the adaptation to the environment and its own body. Despite its simple components and the simple survival criterion, it is difficult to control and predict the robot's behavior, due to non-linearities and feedback connections exploited for optimal navigation and obstacle avoidance.[74]

Ultimately Karnow concludes that "Robot autonomy of course exists across a spectrum."[75]

In increasing degrees of autonomy, robots rearrange logical or physical modules to solve the assigned task and create modules from smaller units, programming themselves.[76]

* * *

The second part of this article, including the bibliography, will be published in an upcoming issue of *The Journal of Robotics, Artificial Intelligence & Law*.

Notes

* Katherine D. Sheriff is a technology associate at Davis Wright Tremaine LLP, devoting her legal practice to identifying areas of opportunity and potential challenges in emerging technology sectors, particularly in the dynamic fields of autonomous vehicles and artificial intelligence. She began writing this article in 2015 while at Emory Law School and has updated it since then to the point where it prioritizes consumer education at the industry level and standardized nomenclature across jurisdictions. Ms. Sheriff may be reached at katherinesheriff@dwt.com.

1. Ugo Pagallo, *The Laws of Robots: Crimes, Contracts, and Torts, Law Governance and Technology Series* 10 (Springer 2013).
2. *Id.* at 114-46.
3. *Id.* at 79-114.
4. *Id.* at 45-78.
5. Mark Riedl, *Law and Technology Visiting Lecture*, Emory University School of Law (2015), https://eilab.gatech.edu/mark-riedl (accessed Nov. 29, 2020).
6. Mark Riedl, *The Lovelace 2.0 Test of Artificial Creativity and Intelligence*, Georgia Tech School of Interactive Computing and Director of Entertainment Intelligence Lab (2015), https://arxiv.org/abs/1410.6142 (accessed Nov. 29, 2020).
7. Mark Goldfeder, *Law and Technology*, Emory University School of Law (2015).
8. Lisa Vertinsky, *The Role of Patents in International Public Health*, Emory University School of Law (2015).
9. Katherine D. Sheriff, *Professional Liability After Quantum Leaps in Technology: The Advent of Autonomous Vehicles and Technology's Uncertain Fit Within Existing Tort Law* (2014).
10. Lisa Vertinsky, *The Role of Patents in International Public Health*, Emory University School of Law (2015).
11. Cadie Thompson, "The Real Problem with Artificial Intelligence" *Tech Insider* (Sept. 17, 2015), https://www.businessinsider.com/autonomous-artificial-intelligence-is-the-real-threat-2015-9 (accessed Nov. 28, 2020).
12. James Niccolai, "Google Says It's 'Rethinking Everything' Around Machine Learning," *PC World* (Oct. 23, 2015), https://www.pcworld.com/article/2996620/google-reports-strong-profit-says-its-rethinking-everything-around-machine-learning.html (accessed Nov. 26, 2020).
13. Stuart J. Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach* (Prentice Hall 1995) vii.
14. M.I. Jordan and T.M. Mitchell, "Machine Learning: Trends, Perspectives and Prospects" 349 Science 255 (2015).
15. Dave Smith, "Technology's Next Big Frontier Is Teaching Machines to Learn from Their Own Mistakes," *Tech Insider* (Oct. 23, 2015), http://www.techinsider.io/what-is-machine-learning-2015-10 (accessed Nov. 26, 2020).

16. Jordan and Mitchell, "Machine Learning" 255.
17. Russell and Norvig, *Artificial Intelligence* 479.
18. *Id.* at 419.
19. *Id.* at 479.
20. *Id.* at 27.
21. *Id.* at 31.
22. *Id.*
23. *Id.* at 5 Figure 11.
24. *Id.* at 394 (emphasis in original).
25. *Id.* (emphasis in original).
26. *Id.* at 456.
27. *Id.* at 436-437.
28. *Id.* at 508 Figure 17.8.
29. *Id.* at 485-486.
30. *Id.* at 486.
31. *Id.* at 486, 419 Figure 14.1.
32. Jordan and Mitchell, "Machine Learning" 255.
33. Russell and Norvig, *Artificial Intelligence* 525-529.
34. *Id.* at 527.
35. *Id.* at 527 Figure 18.3.
36. *Id.* at 531-534.
37. *Id.* at 537 Figure 18.7.
38. *Id.* at 552-557.
39. *Id.* at 552.
40. *Id.* at 553.
41. *Id.* at 556 Figure 18.17.
42. Jillian D'Onfro, "Google Is 'Re-thinking' All of Its Products to Include Machine Learning," *Business Insider* (Oct. 22, 2015), http://www.businessinsider.com/google-on-machine-learning-2015-10 (accessed Nov. 26, 2020).
43. Niccolai, "Google Says It's 'Rethinking Everything' Around Machine Learning."
44. Bernard Marr, "5 Ways Machine Learning Is Reshaping Our World," *Forbes* (Oct. 22, 2015), https://www.forbes.com/sites/bernardmarr/2015/10/22/5-ways-machine-learning-is-reshaping-our-world/?sh=49d7b316641a (accessed Nov. 26, 2020).
45. Valerie Bonnelle and others, "Individual Differences in Premotor Brain Systems Underlie Behavioral Apathy" (2015) Cerebral Cortex 1.
46. Bonnelle and others, "Individual Differences" 1.
47. *Id.*
48. *Id.*
49. *Id.*
50. *Id. See also*, Dean Mobbs et al., *From Threat to Fear: The Neural Organization of Defensive Fear Systems in Humans*, 29 J. Neurosci. 12236, 12240-12241 (2009).

51. *Id.*

52. R. Legenstein and others, "A Reward-Modulated Hebbian Learning Rule Can Explain Experimentally Observed Network Reorganization in a Brain Control Task" (2010) 30 J Neuroscience 8400.

53. Gregor M. Hoerzer and others, "Emergence of Complex Computational Structures" (2014) Cerebral Cortex 678.

54. Alan Turing, "Computing Machinery and Intelligence" (1950) 59 Mind 433; John R. Searle, "Minds, Brains and Programs" (1980) 3 Behavioral and Brain Sciences 416; Lawrence Solum, "Legal Personhood for Artificial Intelligence" (1992) 70 NC L Rev 1231.

55. Robert I. Damper, "The logic of Searle's Chinese Room Argument" (2006) 16 Mind 164-165.

56. Solum, "Legal Personhood for Artificial Intelligence" 1273.

57. Damper, "The Logic of Searle's Chinese Room Argument" 165.

58. Solum, "Legal Personhood for Artificial Intelligence" 1273.

59. Searle, "Minds, Brains and Programs" 422.

60. *Id.*

61. Solum, "Legal Personhood for Artificial Intelligence" 1273.

62. Tom Allen and Robin Widdison, "Can Computers Make Contracts?" (1996) 9 Harv JL & Tech 25, 39; Günter Teubner, *Rights of Non-Humans? Electronic Agents and Animals as New Actors in Politics and Law*, Max Weber Lecture at the European University Institute of Fiesole, Italy (Jan. 17, 2007).

63. Daniel Dennett, *The Intentional Stance* (MIT Press 1987) 1.

64. Allen and Widdison, "Can Computers Make Contracts?" 30.

65. *Id.*

66. Pagallo, *The Laws of Robots* 103-104.

67. Thompson, "The Real Problem with Artificial Intelligence."

68. *Id.*

69. Giulio Prisco, Press Release: "Susan Schneider on Machine Consciousness at the Upcoming Terasem Colloquium in Second Life" (*Turing Church*, Dec. 4, 2015), http://turingchurch.com/2015/12/04/susan-schneider-on-machine-consciousness-at-the-upcoming-terasem-colloquium-in-second-life/ (accessed Nov. 27, 2020).

70. Prisco, "Susan Schneider on Machine Consciousness."

71. Robert M. French, "Subcognition and the Limits of the Turing Test," in Stuart M. Shieber ed., *The Turing Test: Verbal Behavior as the Hallmark of Intelligence* (MIT Press 2004) 193-194.

72. French "Subcognition and the Limits of the Turing Test" 193-194.

73. Curtis E.A. Karnow, "The Application of Traditional Tort Theory to Embodied Machine Intelligence," in Ryan Calo and others eds., *Robot Law* (Edward Elgar Publishing 2016).

74. Karnow, "The Application of Traditional Tort Theory to Embodied Machine Intelligence" 51; D. Floreano and others, *Automatic Creation of an Autonomous Agent: Genetic Evolution of a Neural-Network Driven Robot*

(1994) Proceedings of the Third International Conference on Simulation of Adaptive Behavior: From Animals to Animats 421.

75. Karnow, "The Application of Traditional Tort Theory to Embodied Machine Intelligence" 51.

76. *Id.*; D. Floreano and others, *Automatic Creation of an Autonomous Agent* 421.