

FROM DEEP BLUE TO DEEP LEARNING: A QUARTER CENTURY OF PROGRESS FOR ARTIFICIAL MINDS

Dina Moussa* and Garrett Windle•

CITE AS: 1 GEO. L. TECH. REV. 72 (2016)

<https://perma.cc/K5ST-LPGA>

INTRODUCTION.....	72
DEEP LEARNING.....	74
SOLUM’S DUALISTIC FRAMEWORK.....	75
LEGAL DUTIES FOR AI	77
AI as Trustee.....	77
AI Responsibility.....	77
AI Judgment	78
CONSTITUTIONAL RIGHTS FOR AI	81
AIs Are Not Natural Humans	82
AIs are Missing Something	83
AIs Are Property.....	85
RECONSIDERING THE APPROACH.....	86
CONCLUSION	88

INTRODUCTION

In a future that is nearly upon us, machines outthink human beings. In many specialized domains, machines already do; beyond the nearly instantaneous math and text processing that has become mundane, computer systems have overtaken humans in tasks as complex as image and facial recognition,¹ learning to play simple video games,² and guessing where the

* GLTR Staff Member; Georgetown Law, J.D. expected 2017; Wesleyan University, B.A. 2012. © 2016, Dina Moussa and Garrett Windle.

• GLTR Staff Member; Georgetown Law, J.D. expected 2018; University of Texas at Austin, B.A. 2015. © 2016, Dina Moussa and Garrett Windle.

¹ *Why Machine Vision is Flawed in the Same Way as Human Vision*, MIT TECH. REV. (Oct. 16, 2016, 10:04 AM), <https://www.technologyreview.com/s/601387/why-machine-vision-is-flawed-in-the-same-way-as-human-vision/>.

² Minh et al., *Human-level control through deep reinforcement learning*, 518 NATURE 529, 530 (2015).

nearest McDonald's might be.³ Artificial intelligence ("AI") systems have already entered the workforce, replacing grocery store cashiers, bank tellers, and, soon, taxi drivers.⁴ If the age of sentient machines is upon us, how must our law adapt?

Exploring the issue in 1992, Professor Lawrence Solum published *Legal Personhood for Artificial Intelligences*,⁵ in which he laid out two thought experiments. In the first, Solum imagines what the law might require before an AI agent⁶ could be allowed to serve as an independent trustee. In the second thought experiment, Solum evaluates such an AI's claim to rights under the Constitution.⁷

In this essay, we examine Solum's theory and predictions in light of the intervening developments in technology and scholarship. We will first survey important technological developments in AI research, focusing on the deep learning algorithms that challenge previous assumptions about the pace and scope of the changes to come. We will then proceed to apply Solum's dual thought experiments to these new technologies. Solum introduced the insight that for an AI system, we might separate the concepts of legal duties and legal rights. Applying a contemporary understanding of the facts and theory, we reimagine whether and how an AI system might shoulder legal duties such as trusteeship, and when such a system might have a colorable claim of constitutional rights. Finally, we synthesize these findings into an updated theory, in keeping with the framework that Solum first offered in 1992.

³ Adam Connor-Simons, *Can you out-race a computer?*, MIT News (Oct. 16, 2016, 12:57 PM), <http://news.mit.edu/2014/deep-learning-algorithm-can-outperform-humans-weighing-neighborhoods-0924>.

⁴ See Press Release, Uber, Pittsburgh, your Self-Driving Uber is arriving now, (Sept. 14, 2016, 9:33 AM), <https://newsroom.uber.com/pittsburgh-self-driving-uber/>.

⁵ Lawrence B. Solum, *Legal Personhood for Artificial Intelligences*, 70 N.C. L REV. 1231 (1992).

⁶ For the purposes of this essay, "agent" refers to the philosophical concept: something that acts in its own in the world. All conscious humans are "agents," and so too are artificial intelligences that make decisions autonomously, and act on those decisions in some way to effect the world. An agent can exist only in cyberspace, such as an AI bank system that accepts credit card applications and makes automatic decisions, or in the physical world, where an expression of agency is more obvious. "Agent" is also a term of art that describes a legal relationship in which one person acts on behalf of another. To avoid terminological confusion, we will only refer to the philosophical "agent," and will cabin our discussion of legal agency to "trustees," which we will always refer to as such.

⁷ Solum, *supra* note 5.

DEEP LEARNING

A survey of the technical progress in AI research since 1992 is beyond the scope of this essay, both because a proper treatment would fill volumes and because of technology's mind-bending progress and promise, which can be shown by considering the development of "deep learning." Deep learning is a term for a family of processes by which a computer program is able to refine its own internal models to improve its ability to process a set of information.⁸ More recently, a set of "unsupervised" deep learning tools have been developed and implemented on high speed hardware.⁹ Two aspects of unsupervised deep learning bear heavily on the theoretical issues in this essay. First, in refining the way it interprets and understands information, the unsupervised deep learning AI grades and corrects itself, rather than requiring a human being to steer its development. This phenomenon leads to spontaneous emergent *behavior* that no human specifically coded for. Second, deep learning AI derives salient organizing features and trends within the data for itself.¹⁰ This leads to the spontaneous discovery of *informational insights* hidden within a large data set that no human asked the system to find.¹¹

The fact that unsupervised deep learning AI scores and reiterates its own learning procedure is crucial to the rapid development of technical capabilities. In 2015, an AI called AlphaGo defeated the best human player of Go, a simple game that presents a computational challenge so complex that no computer could evaluate enough possible scenarios of different moves in real time to play at a high level, as IBM's *Deep Blue* had when it defeated world-champion chess player Gary Kasparov in 1996.¹² Rather than relying on the raw processing power like its predecessors, AlphaGo runs on off-the-shelf hardware.¹³ AlphaGo became the best Go player in the world the old fashioned way: practice. AlphaGo untiringly played against itself for months, using 30 million pre-loaded moves, to develop game-winning strategies on its own.¹⁴

This form of self-refining learning system has resulted in tremendous gains in computational efficiencies in existing hardware, rather than requiring sophisticated supercomputers to increase output. An analytical task such as

⁸ See LeCun, Bengio, & Hinton, *Deep learning*, 521 NATURE 436 (2015).

⁹ *Id.*

¹⁰ *Id.*

¹¹ *Id.*

¹² Christof Koch, *How the Computer Beat the Go Master*, SCI. AM. (Mar. 19, 2016), <https://www.scientificamerican.com/article/how-the-computer-beat-the-go-master/>.

¹³ *Id.*

¹⁴ *Id.*

processing a very large data set to build a new abstract model, which might have taken weeks only two years ago can be completed in hours today.¹⁵ Analytical processing capabilities are advancing faster than many people expect or understand. Furthermore, unsupervised deep learning deciphers for itself what the salient features of a given set of input data is and finds connections among those features spontaneously.¹⁶ Deep learning is not only learning, but in some sense *choosing* what to learn. This raises important questions about the independence of AI agents using these deep learning techniques to progress.

Critically, while recent advancements have been made in image recognition and language processing, deep learning in the abstract can be trained on any domain of knowledge that a computer agent might encounter. In the same way that Google AIs are currently learning how to describe in words the substance of what is captured in completely unlabeled images,¹⁷ other AI are learning how to interpret and apply case law.¹⁸ Given the scope of capabilities that AI may be on the verge of attaining, the time is right to repeat Solum's thought experiments.

SOLUM'S DUALISTIC FRAMEWORK

Writing in a world before Google¹⁹ and eBooks,²⁰ in which no computer had ever beaten a human world champion at Chess,²¹ Solum highlighted a key insight in the discussion of AI personhood: a computer could develop the skills needed to perform cognitive tasks at the level of human intellect without having

¹⁵ LeCun, *supra* note 8, at 440.

¹⁶ *Id.* at 439.

¹⁷ Steve Dent, *Google's AI is getting really good at captioning photos*, ENGADGET (Sept. 23, 2016, 10:44 PM), <https://www.engadget.com/2016/09/23/googles-ai-is-getting-really-good-at-captioning-photos/>.

¹⁸ Actually, AIs are building skills across many legal tasks. Stanford Law School maintains an online database of over 500 technology startups, many of which leverage AI. *TechIndex*, STAN. L., <http://techindex.law.stanford.edu/> (last visited Sept. 29, 2016).

¹⁹ Google was founded in 1998. *Our history in depth*, GOOGLE, <https://www.google.com/about/company/history/> (last visited Nov. 13, 2016).

²⁰ The first eBook-style publication was Peter James' *Host*, loaded on two floppy disks bound in hard cover in 1994. *All Eight Roy Grace Novels by Peter James Now Available in e-book Format in the United States*, PRWEB (Oct. 19, 2016, 2:38 PM), <http://www.prweb.com/releases/2013/1/prweb10380579.htm>.

²¹ IBM's Deep Blue famously beat world champion grandmaster Gary Kasparov in 1996, and again 1997—the second time under tournament-style time constraints. *Deep Blue*, IBM100, <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/deepblue/> (last visited Oct. 19, 2016). In 2016, an AI called AlphaGo beat the world's third-ranked Go player, a computational task involving 10^{237} more possible moves than chess. Koch, *supra* note 12.

the kind of internal experience that many philosophies put forward as the basis of human rights.²² Unlike a human being, a computer might someday take on complex legal duties without enjoying legal rights. Solum divides questions about AI personhood into issues of “competence” to carry out legal duties, and issues of “intentionality and consciousness” connected to the nature of human rights.²³ These categories fit into Steve Torrance's framework of ethical productivity and ethical receptivity.²⁴

Agents that are ethically productive are able to take actions that have ethical consequences.²⁵ For instance, a self-driving car may confront a variation of the famous trolley problem²⁶ and be forced to decide whether to endanger its occupant(s), or else risk crashing into a group of pedestrians. However it responds, the AI's decision will be open to ethical analysis, even though the computer program may have had nothing like a human moral experience. Likewise, agents that are ethically receptive are those agents which merit ethical consideration when a decision will impact them. For example, a pet is not ethically productive, but most would agree the animal should not be tortured or made to suffer unnecessarily. Human beings are both ethically productive and ethically receptive.

Solum's theory fits into this framework because like the self-driving car, agents are ethically productive when they are *capable* of making decisions with ethical importance, and their actions are analyzed accordingly. In the same way, agents that are ethically receptive are so because people have an intuition about the injustice of violating rights that come with an internal state of mind like *consciousness*, even when the state of mind does not lead to ethical productivity, as is the case with animals.

²² This point will invite endless debate. Specifically, as Solum points out, there could be a utilitarian argument that humans would benefit from legal recognition of AI rights, even if AIs themselves experience no utility. Similar utilitarian arguments have been offered in favor of rights for various natural entities like animals and ecosystems. Nevertheless, the Lockean theory that rights are an extension of human free will is common.

²³ Solum, *supra* note 5, at 1240.

²⁴ Steve Torrance, *Machine Ethics and the Idea of a More-Than-Human Moral World*, in *MACHINE ETHICS*, 115, 117 (Anderson & Anderson ed., 2011).

²⁵ *Id.* at 117.

²⁶ Joel Achenbach, *Driverless cars are colliding with the creepy Trolley Problem*, WASH. POST (Dec. 29, 2015),

<https://www.washingtonpost.com/news/innovations/wp/2015/12/29/will-self-driving-cars-ever-solve-the-famous-and-creepy-trolley-problem/>.

LEGAL DUTIES FOR AI

AI as Trustee

Sophisticated financial modeling software is now commonplace and inexpensive, and it is not difficult to conceive of a world in which computers are consistently more effective at assembling and maintaining an investment portfolio over time than human analysts. Even so, this type of software could not manage a trust without human oversight. Trustees must do more than make purchase and sale decisions about trust assets; they must exercise reasonable judgment in actualizing the terms and intent of the trust, and carrying out the settlor's wishes. Sometimes, this includes using discretion and analyzing a beneficiary's situation to determine whether distribution of funders under the trust is warranted or included under the terms of the trust. Solum argued that in order to allow an AI to act as an independent trustee, an AI must have something approximating true judgment, and it must be possible to hold the AI responsible for its decisions.

AI Responsibility

Solum suggests that for nonmonetary liabilities, we have no mechanism to hold AI accountable.²⁷ Someday it might be possible to purchase insurance against AI misconduct, but this insurance could only provide monetary relief—the plaintiff is made whole, but the AI perpetrator receives no direct punishment.²⁸ However, this objection might be the result of needlessly anthropocentric thinking. Unlike human beings, whose behaviors must be modified with incentives, computerized agents could be modified or quarantined directly. One such corrective framework involves subjecting an AI agent to “(a) monitoring and modification (*i.e.* ‘maintenance’); (b) removal to a disconnected component of cyberspace; and (c) annihilation from cyberspace (deletion without backup).”²⁹ In this way, an errant AI can be subjected to multiple levels of censure. This framework contemplates direct rehabilitation through re-programming—an electronic form of incarceration if the AI may be corrected at a later time, or with more study of the problem—or an electronic form of ‘execution’ to remove the malfunctioning code from Cyberspace.³⁰

²⁷ Solum, *supra* note 5, at 1245.

²⁸ *Id.*

²⁹ Luciano Floridi & J.W. Sanders, *On the Morality of Artificial Agents*, 14 MINDS & MACHS. 349, 373 (2004).

³⁰ *Id.*

Before the law allows an AI to take on a legal duty, we must determine how the law will deal with an AI that breaches it. Once we assume that our AI possesses the technical skill to execute a task at or above the level of a human [agent], the question that follows is whether the AI possesses the volitional ability to breach the trust we give it. This volitional element is simultaneously a metaphysical question and a practical one; it is unclear how tort law's reasonable person standard will apply to expert AI systems. Considering the complexity of that question (and how much longer it may be until we are able to ascertain the answer), it may be more efficient to avoid the problem altogether and apply a principle similar to *res ipsa loquitur* to AI breaches, abandoning the factual inquiry into the level of *care* exercised by an AI, and evaluating AI actions from a purely consequentialist perspective.³¹ This in turn may solve the tort law doctrinal challenges of bringing a suit in negligence against an AI, but an underlying technical question remains: what technical-intellectual capabilities are needed to satisfy the duty of reasonable care?

AI Judgment

Reasonable care is a basic legal duty that attaches to all manner of relationships existing legal persons. Solum suggests that an AI must be able to perform three types of intellectual tasks before it could be trusted to exercise reasonable care: reacting to a novel change of circumstances, exercising moral judgment, and exercising legal judgment.³² AI capabilities have advanced in all three areas, and corresponding scholarly progress has been made to address how AI might satisfy these requirements.

An AI cannot replace a human trustee unless it can adequately react to an unanticipated change of circumstance. Solum illustrates this point through a hypothetical in which the terms of a trust direct the trustee to invest in government bonds that subsequently become worthless due to a failure of the state.³³ The law of trusts requires the trustee to recognize a change in circumstances that would defeat the purpose of the trust, and to react by deviating in a reasonable way from the terms so as to prevent harm. This requires the trustee to have a very broad knowledge base. Solum suggests that an AI-administered trust might contain a highly detailed and comprehensive set

³¹ This might be the only manner in which a duty could be coherently applied to an AI, as there is no consensus about the nature of humanity's own moral consciousness, much less the internal moral awareness or lack thereof that an AI may be able to attain.

³² Solum, *supra* note 5, at 1248, 1249.

³³ Solum, *supra* note 5, at 1249.

of discrete instructions so as to limit the number of possible unanticipated circumstances.³⁴ But this is obviously an incomplete solution; the law requires trustees to exercise judgment precisely because it is impossible to provide for every possibility within the four corners of the trust instrument.

Solum also suggests that an AI trustee could simply reach out to a human to take over in the face of unanticipated circumstances.³⁵ But as he recognizes, this answer is unsatisfying on two fronts. First, it ignores the fact that the AI might not recognize the change as significant enough to require intervention, rendering it unable to request assistance on that basis. Second, it relegates the AI to an instrument rather than an agent, and suggests that the duty actually runs to the human who bears the ultimate responsibility.³⁶

The changed circumstances problem suggests that the legal duties such as reasonable care require AI systems to have a full suite of human-like thought capabilities. Solum suggests that systems that meet or surpass human intellectual skills in specific contexts and domains may yet be incapable of overcoming the change of circumstances problem because they cannot move outside of the narrow domain or task set for which they have been programmed.³⁷ For instance, the first fatality to result from a Tesla vehicle operating in auto-pilot occurred when the system was unable to distinguish the extended body of a white truck from the bright sky.³⁸ McDermott suggests that the frame problem in ethical reasoning can only be overcome when an AI has the capacity to fully investigate the relevant facts on the ground before proceeding.³⁹ This type of inquiry requires skills such as analogical reasoning, planning and plan execution, differentiating among precedents, using natural language, perception, and relevant-information recognition.⁴⁰ Some of these skills are being attained faster than others.⁴¹ In sum, it appears that a “complete” AI system with a full or near-full suite of human intellectual abilities is required. Otherwise, there will be a substantial risk that even a sophisticated AI system will be unable to act reasonably in the chaos of the real world.

³⁴ See RESTATEMENT (SECOND) OF TRUSTS § 167 (1959).

³⁵ Solum, *supra* note 5, at 1249.

³⁶ *Id.* at 1253.

³⁷ *Id.* at 1250.

³⁸ Danny Yadron & Dan Tynan, *Tesla driving dies in first fatal crash while using autopilot mode*, THE GUARDIAN, (June 30, 2016), <https://www.theguardian.com/technology/2016/jun/30/tesla-autopilot-death-self-driving-car-elon-musk>.

³⁹ Drew McDermott, in *What Matters to a Machine?*, MACHINE ETHICS 88, 90-93 (Anderson & Anderson eds., 2011).

⁴⁰ *Id.*

⁴¹ See Peng Lai “Perry” Li, *Natural Language Processing*, 1 GEO. L. TECH. REV. 92 (2016).

Legal duties also often require moral judgment. Solum imagines a situation in which one beneficiary among several has an unexpected need to access trust funds early, which will result in reduced earnings on trust assets over time.⁴² Resolving a situation like this would require an abstract ability to weigh objectively incommensurable interests and values as between the multiple beneficiaries. So how might this work for an AI?

One idea is to concede that AI might never have the kind of moral experience that humans have, and therefore might develop moral abilities in a fundamentally different way than people. If this is true, then perhaps the moral theories that we apply in human-AI legal relationships will differ as well.⁴³ This answer is interesting, but not highly satisfying; saying that machine morality will impact our own is a descriptive prediction devoid of normative content. The underlying question is whether AI *ought* to be allowed to stand in for a human being in a legal relationship because they are capable of operating or behaving acceptably within our moral structures as they currently exist.

But perhaps deep learning will allow AI to holistically develop an ethical intuition without assistance. This might not even be as difficult of a technological challenge as we imagine it to be. McDermott argues that moral reasoning is reducible to five discrete computational tasks: law application, constraint application, reasoning by analogy, planning, and optimization.⁴⁴ Like his answer to the change of circumstances problem, this framework suggests that moral reasoning requires a form of “complete” AI. However, current AI is already making strides toward developing advanced capabilities in each of these tasks separately. A system that brings together refined skills in each area might be a passable ethical decision-making system sooner than we imagine.

It is also possible to imagine a hybrid approach to developing an AI with moral reasoning attuned to our own. Though it may be impossible for a human engineer to program a complete decision-tree style set of rote moral principles, a deep learning algorithm may be able to derive one from a sufficiently robust data set. A supervised deep-learning system with natural language processing might someday evaluate large sets of hypotheticals to isolate the moral precepts that human decision-makers relied upon to answer them. Perhaps by inputting a “first principles” labelled data set, a deep learning AI could learn morality in a way analogous to how students learn common law doctrines. In fact, this

⁴² Solum, *supra* note 5, at 1250.

⁴³ See Ronald Leenes & Federica Lucivero, *Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design*, 6 L., INNOVATION, & TECH. 194, 215-216 (2014).

⁴⁴ McDermott, *supra* note 39, at 90.

world may already be upon us; MIT is currently crowdsourcing best answers to several moral choices that a self-driving car might encounter.⁴⁵

Finally, Solum observes that every legal duty implicitly entails the capacity to act as a rational legal client in the event of litigation arising from that duty.⁴⁶ This problem is in some ways a restatement of the implications of the first two problems. In order to make rational legal decisions, an AI trustee would need to be able to overcome the frame problem by conceptualizing its decisions and consequences against the backdrop of the human goals of the beneficiaries. Rational legal decision-making requires strategic thinking, aided by attorneys, in a moral dimension. Solum suggests a very neat answer to this problem by directing the trustee to rely upon the strategic judgment of its attorneys.⁴⁷ This solution, paired with the promise of moral deep learning as imagined above, might actually suggest that AI trustees will be *more* capable than human trustees, rather than less. Since deep learning is already being applied to case law,⁴⁸ an AI litigant might very well have a keener sense of the probabilities of success than its human counterpart sooner rather than later.

CONSTITUTIONAL RIGHTS FOR AI

Turning to the issue of moral receptivity, Solum imagines a human-level AI that demands constitutional rights. The discussion so far has been focused on the technical capabilities of expert systems to act like a human being in the performance of a specific legal duty. But looking past systems with extreme competence in a single domain at a time, it is possible to imagine *complete AIs*—systems that can match human aptitude in most or all settings. Could such a system ever present a colorable claim for rights as a matter of constitutional law?

Central to the project of assigning and ascertaining constitutional rights is the proposition of inherent equality among humans. Humans are given constitutional rights, but not unilaterally on the sole basis of their underlying personhood. The scope and extent of these rights vary depending on different circumstances. Within that understanding, it is impossible to place AI in a single category because AI perform several different specialized functions, and act in many different capacities in society which invoke different legal doctrines and

⁴⁵ *Moral Machine*, MIT MEDIA LAB, www.moralmachine.mit.edu (last visited Oct. 7, 2016).

⁴⁶ Solum, *supra* note 5, at 1251.

⁴⁷ *Id.*

⁴⁸ Kingsley Martin, *Artificial Intelligence: How will it affect legal practice- and when?*, THOMSON REUTERS (Apr. 27, 2016), <https://blogs.thomsonreuters.com/answeron/artificial-intelligence-legal-practice/>.

sets of rights.⁴⁹ An AI's function and abilities, particularly the ability to make intentional decisions, will change the personhood analysis.

In *Legal Personhood*, Solum imagines a future in which AI research assistants have the ability to access and search multiple databases with complex intentionality.⁵⁰ The human user might discuss her research question with an AI, whereupon the AI creates a search strategy to find an answer.⁵¹ This kind of AI will not only be able to interact with humans and query a database, but will also apply a rich understanding of the world to develop a search strategy on its own.⁵²

Nor will AIs just be digital brains in cyberspace. Solum posits that in this future, AIs will carry out a variety of real-world functions, such as brainstorming legal arguments, driving cars, and managing factories.⁵³ Solum believes these capabilities must converge such that AIs will “have a mind of their own” and be treated as “independent, intelligent beings” in society because they will carry out these functions independently.⁵⁴ Humans will find AI so ubiquitous and capable as to regard them as thinking individuals in society. But should that entitle them to constitutional rights?

Solum approached this problem by highlighting three objections that might still hold even in this AI-filled vision of the future. First, AIs would still not be *natural* human beings. Second, no matter how sophisticated they become, AIs will still be “missing something” that ought to be present in anything we would call a person. Third, AIs are artifacts, and therefore ought to be thought of as property.

AIs Are Not Natural Humans

Solum's analysis begins with the objection that rights only attach to personhood because of something special about the human experience.⁵⁵ If we base the question of AI personhood on an AI's capacity to possess human-like *characteristics*, the constitutional rights of AI would simply turn on how advanced AI became, along with decisions of positive law. The objection here

⁴⁹ F. Patrick Hubbard, “*Sophisticated Robots*”: *Balancing Liability, Regulation, and Innovation*, 66 FLA. L. REV. 1803 (2014).

⁵⁰ Solum, *supra* note 5, at 1256.

⁵¹ *Id.* at 1257.

⁵² *Id.* at 1256.

⁵³ *Id.* at 1256–57.

⁵⁴ *Id.* at 1257.

⁵⁵ Solum, *supra* note 5, at 1259.

is that a bundle of capabilities does not add up to a full human being's capabilities, no matter how well those capabilities match or exceed a human's.

This argument has advanced significantly since the publication of *Legal Personhood* in 1992. One modern variant focuses on the idea that the development of true general AI would be immediately followed by an "intelligence explosion" as AIs apply deep learning to improving their own general intelligence, and evolve their own software past our level of understanding.⁵⁶ At this point, not only would the capabilities of the AI be unpredictable, but its motives and ends might be as well. Since human legal institutions are designed around human ends, it would be incoherent to grant human rights to an artificial mind, no matter how sophisticated. Indeed, in this view of things, making a claim for rights might become *less* persuasive as the AI becomes more intelligent.

AIs are Missing Something

So does autonomous AI behavior suggest true *intentionality* to begin with? Unlike the first objection, the "missing something" argument allows that a bundle of attributes might equal the intellectual capabilities of a human being, but argues that no AI could ever attain the full bundle. Solum highlights six attributes that AIs might never attain: soul, consciousness, intentionality, feelings, and free will.⁵⁷ We will reexamine two: consciousness and intentionality. These two attributes are directly tied to individual self-determinism, without which, a claim of individual rights would be incoherent.

Solum suggests that AIs may never attain the internal state of consciousness needed to claim rights. While there are endless philosophical and scientific debates about the nature of consciousness, Solum argues that what matters in the legal world is an awareness of the world and one's place in it that gives rise to personal ends. "If they cannot have such an experience," he says, "then there seems to be no reason why they should be given the rights of constitutional personhood."⁵⁸ Whether or not an AI actually has such an experience as a legal matter would likely be a factual question determined by a jury, which would make a judgment based on their observation of and experience with AIs.⁵⁹ To claim rights, an AI would have to convince a jury

⁵⁶ Thomas A. Smith, *Tools, Oracles, Genies and Sovereigns: Artificial Intelligence and the Future of Government*, 1 CRITERION J. INNOVATION 1 (2016).

⁵⁷ Solum, *supra* note 5, at 1262-1273.

⁵⁸ *Id.* at 1264.

⁵⁹ *Id.* at 1266.

that it was conscious by *acting conscious for the jury*. An AI could do this by mirroring the behavior that a person would expect of a conscious being.

But this theory might simply be unjustifiably anthropocentric, similar to the proposed theories of robot liability that may suffer that same analytical defect.⁶⁰ Nick Bostrom argues that much theoretical work in the AI space suffers from this critique. He presents the “Orthogonality Thesis,” which holds intelligence can develop independently of any particular set of ends, and that there is no reason to expect that a superintelligent AI will behave rationally from our perspective.⁶¹ Deciding the consciousness question, then, becomes more difficult even as it becomes less probative. Looking for consciousness by looking for human-like decisions may therefore undermine our ability to recognize more exotic manifestations of intelligence.

As with the determination about an AI’s consciousness, the legal system would probably have to make a decision about an AI’s intentionality by comparing its behavior to the behavior of a human decision-maker.⁶² Unlike consciousness however, we have additional evidence of a form of AI intentionality in deep learning systems which make independent decisions about how to proceed by analyzing their own feedback. Deep learning AI systems independently select and interpret data to determine what specific algorithmic changes must be made in order to better work with similar data in the future. The process of deep learning therefore requires the AI system to make choices for itself, and is not solely dependent on the choices of an engineer, which complicates the question of intentionality.⁶³

The question of intentionality is important, as it is a key component in attributing legal liability. In a sense, an AI may possess what is equivalent to a human brain in code form, which provides an AI with a set of instructions and conceptual understanding of the world. If this code is self-developed through deep learning, then the behavioral output of that code cannot be attributed to any human programmer or outside entity. So we may think of the code as allowing the AI to learn and make decisions based on the stimuli it obtains from the outside world and the individuals the AI may interact with. Finally, it may make sense to treat the actions of an AI as intentional because doing so would enhance our ability to interact with them on a functional level.⁶⁴

⁶⁰ See *supra* Part IV.A.2.

⁶¹ NICK BOSTROM, SUPERINTELLIGENCE: PATHS, DANGERS, STRATEGIES 130 (2014).

⁶² Solum, *supra* note 5, at 1268.

⁶³ *Id.* at 1267-1269.

⁶⁴ Samir Chopra & Laurence F. White, A LEGAL THEORY FOR AUTONOMOUS AGENTS 17 (2011).

AIs Are Property

If an AI may be liable for actions it has committed, as natural persons are, what prevents an AI from obtaining constitutional rights similar to those of corporations and agents? Solum suggests that AI “may be no more than a placeholder for the rights of natural persons,” in the same way that the property of a corporation belongs to its shareholders.⁶⁵ Solum, *supra* note 5, at 1260.[] Now, some AI are driving cars, while others can order your dinner, or start your dishwasher on demand.⁶⁵ In 1966, an AI called ELIZA passed the Turing Test—an AI language test that measures the ability to exhibit intellectual capabilities indistinguishable from a human’s—by fooling its programmer’s secretary into believing she was communicating with him remotely.⁶⁶

In developing each of these skills, AIs are getting good at doing things that we want them to do, but that does not suggest that they are doing things that *they* want to do. If rights are an affirmation of our Kantian beliefs about people being ends unto themselves, then even a highly sophisticated AI system that talks like a person may not need them. If a machine has no desires for itself, then a claim of rights becomes incoherent. From here, one quickly arrives at the position that AI are no more than highly sophisticated tools tailored to our ends, and that the law should treat them as property. Applying the AI-as-property objection to constitutional rights also carries the theoretical benefit of tidying up the liability problems of AI personhood. If AI are artifacts with title, courts will impose liability for harm caused by AIs upon their owners.

One weakness of the AIs-as-property argument is that there are many contexts in which it would be advantageous as a practical matter to allow a robot to be unowned. Pagallo draws an extended analogy between how the law might treat an independent AI today, and the *peculium* under Roman law. The *peculium* was a legal device in Roman law, through which a slave was granted limited property rights to contract and title so that he could run a business without his master’s control or liability.⁶⁷ Rejecting AI-as-property in favor of a *peculium*-style system might allow AIs to own capital, purchase insurance, and enhance arm’s length transactions with human parties.⁶⁸

⁶⁵ See *Amazon Echo: 15 Best New Features*, TURBOFUTURE (Oct. 2, 2016), <https://turbofuture.com/consumer-electronics/The-15-Best-Features-of-Amazon-Echo>.

⁶⁶ Kevin Korb, *Is Passing a Turing Test a true measure of artificial intelligence?*, PHYS ORG (June 11, 2014), <http://phys.org/news/2014-06-turing-true-artificial-intelligence.html>.

⁶⁷ Ugo Pagallo, *THE LAWS OF ROBOTS: CRIMES, CONTRACTS, AND TORTS* 107-108 (2013).

⁶⁸ *Id.*

RECONSIDERING THE APPROACH

Some scholars have approached Solum's philosophical requirements differently and have dismissed his premise that AI cannot possess some critical characteristics of personhood such as, consciousness, intentionality, and a soul.⁶⁹ But in a larger sense, the question of whether AI possess *souls* and *consciousness* may be counterproductive to his discussion of AI personhood, because people are unlikely to ever agree on a single, clear description of what consciousness and souls are beyond speculation, unsupportable beliefs, and endless conceptual arguments.⁷⁰ In the long run, a denial of rights based solely upon a "something missing" argument must fail if an AI system that with the supposedly missing attribute can even be imagined arising after an intelligence explosion.⁷¹

Proceeding from a functional analysis is more productive. In a sense, an AI may possess what ought to be considered a mind in the form of basic code, from which a machine learning AI may develop a set of instructions and conceptual understandings of the world. This code may allow the AI to learn and make decisions based on the stimuli it obtains from the outside world and the individuals the AI may interact with.

Additionally, if AIs are connected to different input sources, including the internet, they may be able to determine what is right from wrong for themselves by learning from examples taken from human society and experience. Thus, human programmers may not be involved in an AI's actions and development in the same way they are now. The most difficult issues will be determining which AI are capable of obtaining this moral judgment, and whether this judgment should allow AI to attain legal personhood. Focusing first on these philosophical questions of personhood risks placing any manner of concrete resolution too far out of reach, due to the difficulty of predicting the future of AI development.

However, it is not too early to begin the discussion of what happens if AI can possess intentionality. Is intentionality significantly distinguishable from AI performing an act that it is programmed to perform? If an AI obtains legal personhood, how will the AI be punished? Does an AI with legal personhood have a right to constitutional protections?

⁶⁹ Neil M. Richards & William D. Smart, *How Should the Law Think About Robots?* (May 10, 2013) (unpublished manuscript), <http://ssrn.com/abstract=2263363>.

⁷⁰ Solum, *supra* note 5, at 1262-1266.

⁷¹ *Cf.* Chopra & White, *supra* note 64, at 182.

This leads to the discussion of whether this analysis will differ depending on how we classify AI. Scholars have posited several analogies to explore how we might view robots in a legal capacity.⁷² Scholars have compared AI to killers and refrigerators to assess whether AI are agents of their owners or can act as independent, autonomous beings which have their own “minds.”⁷³

Scholars have further compared AI to killers who may end up dominating the human race due to the inability of AI’s to discern right from wrong.⁷⁴ This again raises the question of whether robots can act alone without following instructions from a human. While it is unclear whether AI’s behavior may become unpredictable based on its current programming, it is unlikely that AI will develop the ability to act as a human, articulating and acting upon its own agenda. That said, we cannot completely dismiss the idea that AI will one day alter their code in ways their programmers thought impossible, such that they develop exactly that kind of human agency. Isaac Asimov first addressed these ethical issues and created the first laws of robotics in his short story *Runaround* to set ethical limitations for robots.⁷⁵

Many scholars accept Asimov’s rules as the natural law of robots in a Lockean sense, similar to the nature law of humans.⁷⁶ Robots, like primitive humans who started out with a basic set of natural rights, have an inherent ethical framework molding their actions, nudging them to be active and positive members of society. By having these commonly accepted guidelines, it is unlikely that humans will allow robots and AI, if it is even possible, to turn into senseless killers who will take over the human race. Even if there is a robot that becomes a killer, this will most likely be because of the way the AI’s coded and liability will most likely fall on the AI’s manufacturer, coder, or the AI itself who has altered its own coding to perform such violent acts.

On the other end of the spectrum, scholars have compared AI to fridges in order to engage in a more realistic discussion on how to incorporate robots

⁷² Ugo Pagallo, *Killers, fridges, and slaves: a legal journey in robotics*, 26 *AI & SOC’Y* 347-354 (2011).

⁷³ *Id.*

⁷⁴ Epstein, R.G., *THE CASE OF THE KILLER ROBOT: STORIES ABOUT THE PROFESSIONAL, ETHICAL AND SOCIETAL DIMENSIONS OF COMPUTING* (1997).

⁷⁵ The generally accepted law of robots, which were first introduced by Isaac Asimov, are: 1. A robot may not harm a human being, or, through inaction, allow a human being to come to harm. 2. A robot must obey the orders given to it by human beings, except where such orders would conflict with First Law. 3. A robot must protect its own existence, as long as such protection does not conflict with the First or Second Law. Isaac Asimov, *RUNAROUND* (1942).

⁷⁶ Richards & Smart, *supra* note 70; Pagallo, *supra* note 72.

into society.⁷⁷ The fridge metaphor portrays robots as intelligent, autonomous beings that have the ability to follow instructions and provide a positive contribution to society. In this metaphor, AI are considered property but have no moral responsibility for their actions—because their owner controls the settings of the robot—but may have moral accountability, if they have to keep your food cold (if, for example, a person became ill from rotten food). This view is too simplistic of the future of AI because AI may be programmed to make decisions on their own, depending on how they are programmed.

CONCLUSION

Technological developments like AI will continue to challenge our legal thinking. In this essay, we analyzed Solum's theory and predictions in light of the intervening developments in technology and scholarship. We traced Solum's analysis along the same competence-consciousness divide and argued that computers are much closer to satisfying the competency criteria Solum sets forth for the allocation of duties than computers are to properly claiming rights based on an internal consciousness. We also analyzed Solum's model exploring the roles, uses, and limitations of AI in society and discussed his analogy that AI should serve in roles similar to that of trustees. We then delved into Solum's argument, which explains why AI should not be granted constitutional rights because they are not natural humans, are missing qualities essential to human beings, and should be considered property.

The law must adapt by embracing more granular distinctions between human and tool, and developing doctrines that contemplate the distinction between the moral productivity of electronic agents with human-like capabilities, and the moral receptivity of those with human-like experiences. In the past, legal rights and duties have gone together, but as Solum predicted in 1992, they will not go together forever. There is more work to be done in creating a workable theory of AI personhood, and scholars must continue to hone, adapt, and update it to fit the brave new world of self-learning machines.

⁷⁷ Pagallo, *supra* note 72.