Focus

Artificially Intelligent Law

Lawrence B. Solum*

ABSTRACT: This paper explores a series of thought experiments that postulate the existence of "artificially intelligent law." An artificially-intelligent legal system is defined as one with three functional capacities: 1. The system has the capacity to generate legal norms. 2. The system has the capacity to apply the legal norms that it generates. 3. The system has the capacity to use deep learning to modify the legal norms that it generates. The paper then considers the question whether such a system would be desirable as a matter of legitimacy and justice. The core idea of the paper is that the key to the evaluation of artificially intelligent law is to focus on the functional capacities of the system in comparison to comparable human systems, such as regulatory agencies.

KEYWORDS: artificial intelligence; expert legal systems; legitimacy; justice; our robot overlords

1. Introduction: Legal Personhood for Artificial Intelligences

In 1992, "Legal Personhood for Artificial Intelligences" was published.[1] In that article, I suggested a thought experiment that might illuminate the nature of legal and moral personhood. Here is a slightly modified version of the scenario:

Imagine expert systems developing from existing programs that perform some of the component functions of a trustee. For example, the decision to invest in publicly traded stocks is made by a computer program in what is called "program trading," in which the program makes buy or sell decisions based on market conditions. Programs that interact with financial institutions via the Internet are commonplace. It is not difficult to imagine an expert system that combines these functions with a variety of others, in order to automate the tasks performed by the human trustee of a simple trust.

Such a system might evolve in three stages. At stage one, the program aids a human trustee in the administration of a large number of simple trusts. The program invests in publicly traded securities, placing investment orders via modem and electronic mail. The program disburses the funds to the

* *Professor of Law, Georgetown University Law Center. Mail: lsolum@gmail.com.* *The essay has been peer-reviewed by the Scientific Committee.*

[1] LAWRENCE B. SOLUM, *Legal Personhood for Artificial Intelligences*, in *North Carolina Law Review*, 70 N.C. L. Rev. 1231, 1992.

trust beneficiaries via an electronic checking program. Upon being informed of a relevant event, such as the death of a beneficiary, the program follows the instructions of the trust instrument—for example, changing the beneficiary or terminating the trust. The program prepares and electronically files a tax return for the trust. The human trustee operates as do trustees today. The human makes the ultimate decisions on how to invest the funds, although she may rely upon an expert system for advice. She reviews the program's activities to ensure that the terms of the trust instrument are satisfied. But the actual performance of the day-to-day tasks is largely automated, carried out by the program without the need of human intervention.

Stage two involves a greater role for the artificial intelligences (AI). Expert systems are developed that outperform humans as investors in publicly traded securities. Settlors begin to include an instruction that the trustee must follow the advice of the AI when making investment decisions regarding trust assets. Perhaps they do this because experience shows that trusts for which the human overrides the program generally perform less well than those in which the program's decision is treated as final. Moreover, trust administration programs become very proficient at analyzing and implementing the terms of trust instruments. There is little or no reason for the human to check the program for compliance. As a consequence, the role of the human trustee diminishes and the number of trusts that one human can administer increases to the thousands or tens of thousands. The human signs certain documents prepared by the program. She charges a fee for her services, but she devotes little or no time to administering any particular trust.

But there may be times when the human being is called upon to make a decision. For example, suppose the trust is sued. Perhaps a beneficiary claims that the trust has not paid her moneys due. Or imagine that an investment goes sour and a beneficiary sues, claiming that the trustee breached the duty of reasonable care and skill. If such events occur with regularity, the trustee will develop a routine for handling them. She might routinely refer such disputes to her attorneys. In time, the expert system is programmed to handle this sort of task as well. It processes the trustee's correspondence, automatically alerting the trustee when a letter threatening suit is received or process is served. The system prepares a report on the relevant trust from its electronic records and produces a form letter for the trustee's signature to be sent to the trust's attorneys. As the capabilities of the expert system grow, the need for the human trustee to make decisions gradually diminishes.

The third stage begins when a settlor decides to do away with the human. Why? Perhaps the settlor wishes to save the money involved in the human's fee. Perhaps human trustees occasionally succumb to temptation and embezzle trust funds. Perhaps human trustees occasionally insist on overriding the program, with the consequence that bad investments are made or the terms of the trust are unmet. What would happen if a settlor attempted to make the program itself the trustee?[2]

The thought experiment formed the cornerstone of an argument that the decision whether to grant legal personhood to artificial intelligences should turn on the functional capacities of the artificial intelligences. If it were the case that an artificially intelligent program had functional capacities to serve as a trustee, the law ought to confer legal personhood on the artificial intelligence.

In this Essay, I explore a different question. I imagine an artificially intelligent system that performs the functions that we now associate with institutions such as regulatory agencies. What if there were

[2] Id. at 1240–42.

a program with the ability to create legal norms that implement some "intelligible principle" for the governance of some realm of human behavior? What if there were an "artificially intelligent law" with the capacity to create, modify, and implement a regulatory system? And what if this artificially intelligent law did a better job of accomplishing the goals of the regulatory system than humans? In the distant (or not so distant) future in which artificially intelligent laws are possible, should we delegate the functions of a regulatory system to them?

2. Expert Legal Systems and Deep Learning

Expert legal systems already exist, and they have been used to make legal decisions for quite some time. An early example is the "DissoMaster," a software program that was used to calculate various financial awards in connection with the dissolution of a marriage.[3] Systems like this do not have legal authority, but as a practical matter, the awards calculated by expert legal systems may be sticky. Negotiated settlements and judicial decisions may treat the decision made by the expert legal system as dispositive, even if it is not legally binding.

We can easily imagine that an expert legal system like the DissoMaster might actually be incorporated in a legal norm. The legislature could enact a statute that required that financial awards in marital dissolution cases conform to the outputs of the DissoMaster. Such a statute would be the functional equivalent of a statute that spelled out the underlying algorithm as a series of mathematical formulae. The reference in the statute to the DissoMaster would be nothing more than a shorthand way of enacting the formula as law.

Expert legal systems like the DissoMaster do not have the capacities that would be required for a true artificially intelligent law. Expert legal systems are algorithmic: they operate on the basis of a formula, accepting inputs from the human user of the system to produce outputs that are specified by numerical formulas.

Artificially intelligent laws are still science fiction, but the current state of artificial intelligence research is suggestive. Especially important is the idea of "deep learning."[4] Machine-learning programs are self-modifying; they improve themselves.[5] Deep-learning involves the use of neural network architecture to implement machine-learning.

3. The Chinese Intersection

In 2014, I published "Artificial Meaning," which introduced a thought experiment called "the Chinese Intersection."[6] Here is the thought experiment, with some minor modifications:

[3] JAMES A. EIDELMAN, *Software Solutions: A Buyer's Guide to the Best Hard-Working Packages*, in *Family Advocate*, Fam. Advoc., Spring 1989, 42, 46.

[4] DINA MOUSSA & GARRETT WINDLE, *From Deep Blue to Deep Learning: A Quarter Century of Progress for Artificial Minds in Georgetown Law and Technology Review*, 1 Geo. L. Tech. Rev. 72, 74–75, 2016.

[5] DR. TOKIO MATSUZAKI, *Ethical Issues of Artificial Intelligence in Medicine in California Western Law Review*, 55 Cal. W. L. Rev. 255, 260, 2018.

[6] LAWRENCE B. SOLUM, *Artificial Meaning in Washington Law Review*, 89 Wash. L. Rev. 69, 74–76, 2014.

Imagine that you are in the not-so-distant future, looking back at the history of traffic management in Shanghai, an urban megalopolis that we shall imagine became even larger and more densely populated. Somewhere in the heart of this imaginary version of Shanghai there was a supremely complicated intersection, where ten major roads, three highways, six surface light rail lines, and twenty-three pedestrian walkways intersect. Traffic engineers and the city officials tried all the usual tools to manage the intersection. Stoplights, left-turn-only lanes, pedestrian walk/don't-walk signs, speed limits, don't-block-the-box signs with big fines, and so forth. None of it did any good. The "Chinese Intersection" as we shall call it was the source of traffic jams of epic proportions--get stuck at the wrong place at the wrong time, and your five-hour commute lasted for five days. Literally.

At some point, the officials realized that human beings just weren't up to the task of managing traffic in the Chinese Intersection. For every move the human regulators made, frustrated drivers had a countermove—always attempting to circumvent the rules in order to shave a few minutes off the journey through the Chinese Intersection. Economists wrote articles that explained that driver behavior can be modeled using the techniques of game theory. The Chinese Intersection, the economists said, involved multiple Prisoner's Dilemmas. If every driver obeyed the rules, then each individual driver would benefit, but no driver could trust the others, so almost everyone cheated, and the result was chaos. Other economists calculated the effects of the Chinese Intersection on productivity, transportation costs, and real estate values. Everyone knew that the number would be a big one, but when the results of the study were announced, it precipitated a political crisis. The Chinese Intersection imposed annual costs of close to 1.6 trillion Yuan (or approximately 100 billion dollars).

The good news was that the budget for solving the problem became functionally unlimited. Tasks forces met. Entrepreneurs proposed solutions. Politicians called for immediate action. The obvious solution was to eliminate the Chinese Intersection altogether—perhaps a nice park could replace it. But then a simulation was run, and it turned out that this would make the traffic problems in Shanghai far worse. Eventually, a solution was proposed. Traffic in the Chinese Intersection (and eventually all of Shanghai) would be managed by an AI that would operate a system of traffic controls. The system would operate a massive system of programmable lane markers, signs, signals, gates, bumps, broadcasts, robot vehicles, and vehicle removal cranes. The goal was to create a system that would write the traffic laws, inform drivers of their contents, and then enforce the laws by removing non-complying vehicles and collecting on-the-spot fines from the drivers.

And so, a team of programmers and traffic engineers began the process of developing the Shanghai Artificially Intelligent Traffic Authority (SAITA). Early on in the process, the developers realized that SAITA would need the capacity to adapt itself to changes in driver behavior and traffic flow. The programmers produced a system that was capable of changing the traffic code, speed limits, signage, and lane configurations—just about every element of traffic regulation in the city. The system was designed to introduce random variations and run controlled experiments to evaluate the effects of various combinations on traffic patterns. Violations would be detected by an elaborate system of electronic surveillance. Offenders would be identified and immediately removed from traffic by a system of cranes located at key intersections. An automated adjudication system would impose fines for minor offenses and jail terms for more serious violations.

The rollout of the new system was not great. There were bugs and for several days, traffic actually became much worse. SAITA produced a variety of legal texts, signs, and symbols. For example, SAITA periodically revised the punishments associated with traffic code violations--optimizing the level of punishment by modeling the costs and benefits in light of the massive amounts of data it collected on driver behavior. SAITA operated the complex system of traffic signs. In the early days, these signs were simply electronic versions of the familiar signs, but as the system introduced random mutations, the signs began to morph as the system learned new techniques for communicating with drivers. Running a static stop sign resulted in a fine of ¥2000, but if the stop sign was flashing, the fine increased to ¥5000 and if the stop sign was flashing, spinning, and cycling from red to orange and purple, then the fine increased to ¥10,000. The system was even capable of introducing its own vocabulary. When the AI detected a new violation type that had a particularly harmful effect on traffic flow, it could invent a word (represented by a logogram or character) to name the type and add a provision to the traffic code. SAITA could even initiate a public information campaign with entertaining YouTube videos and giant billboards defining the new word and instructing the public about the new law.

Traffic improved noticeably, even in the first few weeks, but as the system began to learn from experience, the effects became dramatic. Lanes were adjusted for traffic flow, the laws were optimized, and the system did a much better job of communicating with drivers than the old human-run system had ever done. Within two years, the traffic flowed smoothly through the Chinese Intersection. Commute times went down, happiness levels went up, and there were even calls to extend SAITA's domain to regulation of things other than traffic.

4. The Concept of an "Artificially Intelligent Law"

The Chinese Intersection thought experiment posits an artificially intelligent law. For the purposes of this Essay, I want to stipulate that the phrase "artificially intelligent law" shall refer to an artificially-intelligent deep-learning system with the following functional capacities:

1. The system has the capacity to generate legal norms. In the Chinese Intersection thought experiment, SAITA has the capacity to generate traffic rules and regulations.
2. The system has the capacity to apply the legal norms that it generates. In the Chinese Intersection thought experiment, SAITA has the capacity to apply the traffic rules it promulgates by citing violators and taking enforcement actions.
3. The system has the capacity to use deep learning to modify the legal norms that it generates in response to information about the effects of its regulations. SAITA has the ability to evaluate the effect of its regulations on traffic flow and then modify those regulations so as to better achieve the goal of reducing wait times at the intersection.

The phrase "artificially intelligent law" could be used in other ways, but for the remainder of this Essay, my use of the phrase will be limited to artificially-intelligent deep-learning systems that satisfy these three conditions.

5. Regulation by Artificial Intelligence

The Chinese Intersection thought experiment was deliberately set in a context in which the deployment of artificially intelligent law would be unlikely to be seen as intensely controversial. Automated law enforcement already exists in the context of traffic laws. For example, many jurisdictions employ camera systems that can read license plates and automatically generate citations for offenses like running a stop light or speeding. As a practical matter, the decision of the automated system approximates finality, as the cost of contesting the decision is likely to exceed the benefits of a successful challenge. If the system is highly accurate, then it is likely to be perceived as legitimate, despite the fact that the policing and adjudicatory functions are performed by an automated system with minimal human intervention.

The Chinese Intersection thought experiment goes beyond policing and adjudication, because SAITA designs, promulgates, and modifies legal norms. Delegating the law-making function to an artificial intelligence is qualitatively different than any current use of artificial intelligence of which I am aware. In the context of the Chinese Intersection thought experiment, this may seem innocuous. The design of traffic rules might well be assigned to obscure technocrats in a real-world legal system. The move from technocrats employing cost-benefit analysis to an artificially intelligent system might even be uncontroversial—especially if the system solved an urgent social problem as it does in the Chinese Intersection thought experiment.

Now imagine the extension of artificially intelligent law to contexts in which the stakes are higher and the design of legal norms implicates issues that have a political, ideological, or moral dimension. Here is a brief list of some possibilities:

- Climate Policy: Imagine that an artificially intelligent system is charged with regulation of greenhouse gas emissions. The system is charged with achieving the optimal strategy for reducing emissions using cost-benefit analysis. The resulting system of regulation and enforcement has distributional effects: some firms and employees are losers and others are winners.
- Prison Regulation: Imagine that an artificially intelligent system is charged with the development of rules and regulations governing the conduct of persons confined in a prison system. The system is charged with developing a system of regulations that will reduce violence, increase the physical and mental health of inmates, and provide optimal rehabilitation and deterrence. The artificial intelligence achieves these goals, but some of its regulations are unpopular with prisoners and others are resisted by the unions representing corrections officers, because they reduce employment levels.
- Terrorism: Imagine an artificially intelligent system charged with the development and implementation of legal norms designed to reduce the levels of terrorist activity. The system implements a pervasive system of surveillance that allows the system to identify and detain persons who are planning or implementing terrorist actions.

The delegation of legal authority to an artificially intelligent legal system would surely be controversial in any of these contexts. Enabling legislation would be resisted, even if it ultimately became law. Civil society groups would howl. In each of the three scenarios above, there would be losers, and undoubtedly their representatives would resist the decision on process grounds.

Delegation of legal authority over climate policy, prisons, and terrorism would be politically contentious, precisely because the design and implementation of legal norms in these areas produce winners and losers and involve issues with ideological and moral dimensions. In the discussion that follows, I will put the political dimension of these controversies to the side and instead focus on the moral dimension—and particularly on the questions of legitimacy and justice that would be implicated by the delegation of legal authority to an artificial intelligence in cases like these.

For the purposes of the discussion that follows, I will assume that *legitimacy* and *justice* are contested concepts. Following Rawls, I will assume that there are disagreements about the nature of legitimacy and justice, leading to competing conceptions of the nature of these concepts.[7] In this Essay, I cannot hope to canvass the possible positions or to defend particular conceptions of legitimacy and justice against their rivals. I will simply assume what I believe to be plausible views about the nature of legitimacy and justice.

6. Delegation and Democracy

The delegation of legal authority to an artificially intelligent legal system would raise questions of legitimacy. The concept of legitimacy is particularly difficult. In my opinion, it is one of the most important ideas for normative legal theory and it is simultaneously undertheorized. The discussion that follows explores three ideas that are rooted in legitimacy:

- Democratic Legitimacy: The idea that legal norms should be authorized by democratic institutions.
- Legitimacy as Transparency: The idea that legal norms should be transparent in the sense that they are accessible to the public.
- Constitutional Role Legitimacy: The idea that officials have legitimate roles, specified by the system of constitutional norms, and that official actions should only be undertaken by officials with constitutional authority with respect to the kind of action.

Each of these three ideas about legitimacy would be implicated by the delegation of legal authority to an artificially intelligent system.

Consider democratic legitimacy. Lawmaking by an artificially intelligent legal system would seem, on the surface, to violate the principle that legal norms ought to be authorized by institutions that possess democratic legitimacy. Once the context of artificially intelligent law is moved from traffic regulation to high-stakes contexts like climate policy, governance of prisons, or terrorism, it is natural to move to science fiction scenarios that involve "our robot overlords."

But the democratic legitimacy of artificially intelligent law is not so simple as our initial reaction to high-stakes scenarios might suggest. What if the delegation of authority to an artificially intelligent system is authorized by a democratic institution, for example, a national legislature? Delegations of authority are a pervasive feature of contemporary legal systems. In the United States, such delegations frequently involve independent regulatory agencies—which are insulated from direct control by the elected executive (the President). I am no expert, but I believe that there is a common perception that regulatory authority in the European Union is exercised by technocrats ("eurocrats") that

[7] JOHN RAWLS, *A Theory of Justice*, Harvard, 1971, 5.

are functionally independent of direct supervision and control by majoritarian democratic institutions.

Of course, it might be argued that independent regulatory agencies and eurocrats are themselves illegitimate, but acting on that idea would be hugely controversial. In the United States, for example, many believe that independent regulatory agencies are a practical necessity given the complexity of regulatory policy and the difficulty of enacting regulatory legislation given the constitutional structure of the legislative process, which requires the acquiescence of a majority of the House of Representatives, a supermajority in the Senate, and either the agreement of the President or, in case of a veto, a two-thirds supermajority in both the House and the Senate. But if democratic legitimacy is satisfied by the initial legislative authorization of the independent regulatory agency, then it would also seem to be satisfied by legislation delegating authority to an artificially intelligent legal system.

Artificially intelligent laws that employ deep learning would raise transparency concerns. Getting at the nature of this problem involves a preliminary step, the examination of the transparency issues that would be raised by rules-based expert systems. Algorithmic expert systems implement rules. The structure of the algorithm may be complex, but it is in principle available for public inspection. Once again, a comparison with delegations to independent regulatory agencies and eurocrats may be illuminating. The complex rules that are implemented by an expert system are not different in kind from the implementation of such rules by humans. As a practical matter, the inner workings of a regulatory agency are no more or less transparent than the workings of an expert system. Moreover, the expert system may have a transparency advantage, because algorithms (if they are designed correctly) are not subject to the manipulation or incompetence that afflicts the implementation of complex rule structures by humans.

Deep learning systems are different, because the decision-making procedures developed by a deep learning system may not be accessible, even to those who designed the system.[8] This is a complex issue—even a cursory discussion is beyond the scope of this Essay, but a few observations can illuminate the nature of the problem. The transparency concern would be severe if an expert legal system were to apply legal norms which were secret because the norms themselves were embedded in a deep learning system. That would be the equivalent of secret laws and would violate the publicity requirement that is widely believed to be an essential element of the rule of law. Artificially intelligent legal systems should be designed so that the norms themselves are public, and the artificially intelligent legal system might even be designed to ensure that the norms it promulgates are comprehensible to those who are governed by them.

Even if the legal norms promulgated by an artificially intelligent legal system were transparent, there is another transparency problem. The process by which the legal norms are designed might itself be opaque in the sense that humans would not be able to inspect and comprehend the method by which the content of the legal norm was specified. But once again, this lack of transparency must be compared to the transparency of the human alternative. Legislative processes, at least in the United States, are famous for their lack of transparency; backroom deals and behind-the-scenes influence-

[8] ANDREW D. SELBST & SOLON BAROCAS, *The Intuitive Appeal of Explainable Machines in Fordham Law Review*, 87 Fordham L. Rev. 1085, 2018.

peddling are widely believed to shape legislation. Similarly, the internal processes of independent regulatory agencies may well be shielded from public view.

It is possible that artificially intelligent legal systems could be designed to provide explanations of their processes. Even if the explanation is only an approximation of the actual process, the resulting transparency might equal or exceed the transparency of human lawmaking. Our evaluation of these issues will be mostly speculation, unless and until artificially intelligent legal systems actually exist.

The final problem of legitimacy arises because of the deep connection between legitimacy and legality.

In constitutional systems that allocate lawmaking authority to particular institutions, the delegation of lawmaking authority to an artificially intelligent legal system would raise a question about about the legitimacy of transferring lawmaking authority from the legislature to the artificial intelligence. In the United States, this question would involve the nondelegation doctrine—the rule that prohibits Congress from delegating its constitutional "legislative power" to another institution (e.g., courts or an executive department). Of course, delegations to independent regulatory agencies are viewed as constitutional so long as the authorizing legislation provides what is called an "intelligible principle" that provides sufficient guidance to the agency. [9]

There is a unifying theme to my analysis of the three problems of legitimacy. In each case, our analysis of the problem requires a functional comparison of the artificially intelligent legal system to the corresponding human systems. If the artificially intelligent system does as well as independent regulatory agencies or eurocrats with respect to responsiveness to democratic institutions, to transparency, and to compliance with an intelligible principle, then the legitimacy objection seems misplaced. Of course, it might well be the case that our current institutions should be reformed to make them more legitimate, but in the absence of such reform, these legitimacy concerns are not unique to artificially intelligent law.

7. Justice and Equity

Consider another possible objection to artificially intelligent law. It might be argued that an artificially intelligent legal system would be insufficiently attentive to concerns of justice. One version of this argument might focus on the distinction between consequentialism and deontology. Thus, it might be objectionable that a self-modifying artificial intelligence employing deep learning techniques would focus on quantitative inputs and cost-benefit analysis (consequentialism) to the exclusion of qualitative judgments about distributional effects and desert (deontology). To begin, we should note that precisely this objection is frequently made to human regulatory processes. More fundamentally, the objection rests on the assumption that an artificially intelligent legal system could not acquire a sense of justice, but that assumption may or may not be correct. If human neural nets can support a sense of justice that is sensitive to considerations of desert and distributive justice, then perhaps deep learning systems acquire the same functional capacities. Indeed, it is at least conceivable that

[9] MEAGHAN DUNIGAN, *The Intelligible Principle: How It Briefly Lived, Why It Died, and Why It Desperately Needs Revival in Today's Administrative State in St. John's Law Review*, 91 St. John's L. Rev. 247, 2017.

an artificially intelligent system would do a better job of incorporating such concerns into the design of legal norms.

Once again, it would be functional capacity that would determine the validity and importance of the objection.

In the case of expert systems that are rule governed, we might well be concerned about the problem of equity (in the Aristotelian sense).[10] No matter how well designed the system of legal rules, the complexity of life can outrun the capacity of the rules to achieve outcomes that are fair. It might be argued that an artificially intelligent legal system would be unduly rigid in the application of general legal rules to particular cases. Such a system, the argument would go, is incapable of doing equity—making an exception to fit the unique circumstances of the particular case. But it is not so clear that this objection to an algorithmic rules-based expert system carries over to an artificially intelligent legal system that employs deep learning. The crucial question concerns functional capacity: if the artificially intelligent law had the functional capacity to do equity that was as good as the capacity of human regulators (which may not be so great), then the objection would disappear.

8. Conclusion: The Value of Thought Experiments

The point of this Essay is not to make the case for artificially intelligent law or to answer objections to this idea. The issue has yet to crystallize, because there is no artificially intelligent system with the functional capacities imagined by the Chinese Intersection thought experiment, much less those that would be required by climate policy, the governance of prisons, or the legal response to terrorism.

Rather, the point of these thought experiments is to help us think today about issues that are likely to creep into the law in bits and pieces as artificial intelligences of various kinds begin to play new and unexpected roles in a variety of contexts. At the same time, the thought experiments may shed new light on old problems. Our concerns about artificially intelligent law might help us to think about the design of human law. Questions about legitimacy and justice are deep. Approaching these questions from a new angle just might help us to question old assumptions and develop novel theories. Or so we can hope.

[10] ROGER A. SHINER, *Aristotle's Theory of Equity in Loyola of Los Angeles Law Review*, 27 Loy. L.A. L. Rev. 1245, 1994.

ISSN 2284-4503