

Article

A Systems and Control Theory Approach for Law and Artificial Intelligence: Demystifying the “Black-Box”

Woodrow Barfield

Whitaker Institute, National University of Ireland, H91 TK33 Galway, Ireland; Jbar5377@gmail.com

Abstract: In this paper, I propose a conceptual framework for law and artificial intelligence (AI) that is based on ideas derived from systems and control theory. The approach considers the relationship between the input to an AI-controlled system and the system’s output, which may affect events in the real-world. The approach aims to add to the current discussion among legal scholars and legislators on how to regulate AI, which focuses primarily on how the output, or external behavior of a system, leads to actions that may implicate the law. The goal of this paper is to show that not only is the systems output an important consideration for law and AI but so too is the relationship between the systems input to its desired output, as mediated through a feedback loop (and other control variables). In this paper, I argue that ideas derived from systems and control theory can be used to provide a conceptual framework to help understand how the law applies to AI, and particularly, to algorithmically based systems.

Keywords: algorithms; artificial intelligence; control theory; systems theory



Citation: Barfield, W. A Systems and Control Theory Approach for Law and Artificial Intelligence: Demystifying the “Black-Box”. *J* **2021**, *4*, 564–576. <https://doi.org/10.3390/j4040041>

Academic Editor: Larisa Ivascu

Received: 7 August 2021

Accepted: 13 September 2021

Published: 27 September 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In a typical system controlled by artificial intelligence (AI), data serve as an input to the system, and algorithms use the data to make decisions that impact events in the real world. Given the “input–output” structure of an AI-controlled system, this paper discusses ideas derived from systems and control theory and how they can be used to describe and evaluate the performance of AI in the context of law. In this paper, control theory is used to represent the formal process in the control structure of a system that includes a feedback loop that functions to assess and respond to discrepancies in the performance of the system [1]. As discussed here, such discrepancies may lead to outcomes that may implicate the law and particularly the law of negligence and strict products liability. Additionally, I define systems theory as a theory that describes and explains the various parts and sub-parts that comprise a system and how they may interact to achieve a desired system goal [2]. For AI and law, I connect systems and control theory in the following way: control theory may be used to shed light on the relationship between system components and can help to explain how feedback between system components serves to control a system. I propose that factors that control a system should be of interest not only to engineers and computer scientists tasked with designing AI systems but also to legal scholars and legislators with an interest in how AI should be regulated. Further, it is often stated that AI-controlled systems can be represented as a “black-box”, which is described as complex and opaque and makes decisions using algorithms that exceed human understanding [3]. Therefore, given that algorithms are a main feature of an AI-controlled system, (1) a discussion of systems and control theory should be useful for legal scholars and legislators with an interest in understanding how AI may affect the performance of a system, and (2) a discussion of systems and control theory should be useful to the legal community in their efforts to regulate current and future versions of AI systems.

For the purposes of this paper, systems exist in several forms, but basically, I dichotomize systems as comprising either a physical entity (e.g., a robot) or non-physical

entity (e.g., a government agency). As background, general systems theory can be traced to the 1940s, when biologist Ludwig von Bertalanffy began the study of complex biological systems [4]. Early systems theory introduced key concepts such as open and closed loop systems and stressed the role and importance of context and environment [5]. In this paper, I argue that such concepts also apply to AI-controlled systems whose performance (or output) as expressed in the real world may lead to challenges to current law.

When litigating a civil or criminal law dispute involving the performance or output of an AI-controlled system, a consideration of the relationship between a system's input to its desired output is not only relevant to a physically embodied AI entity (such as an automated guided vehicle) but also to other types of AI systems whose performance may impact events in the real world. For example, when Internet websites are used for online shopping, algorithms not only direct a transaction but make decisions that impact the consumer's purchasing choices [6]. Here, an algorithm may consider various input variables—online browsing behavior, purchase histories, magazine subscriptions, geolocation and movement, educational achievement, salary, race, ethnicity, family relationships, and more—to predict the likelihood someone with a particular constellation of characteristics will engage in certain behaviors [7]. With algorithmic-based systems, how should the law respond if the predictions made by algorithms are inaccurate, biased, or wrong? What areas of law are implicated, and does knowledge of how the system made its decisions or how the system was internally controlled have relevance for litigating a dispute? Essentially the question addressed in this paper is whether the law need only be concerned with the output or external behavior of a system controlled by AI or should the law also be concerned with the input (or data) to a system, how it is processed by algorithms, and how the internal structure of the system leads to certain outcomes. In this paper, I argue that taking the latter approach is necessary: (1) for developing an understanding of how AI, and particularly systems with algorithms, impact the law, and (2) for developing a conceptual framework for regulating law and AI.

When analyzing a system, it should be noted that many AI-controlled systems can be described as a closed-loop feedback system that operates by comparing the desired system output with an input value provided to the system [2]. To illustrate this concept, consider a mobile robot that uses data as input and makes movement corrections (mediated through a feedback loop) as it navigates the environment (output); this is a highly simplified but classic example of a closed-loop feedback system. If a person is injured as a result of the robot's actions, given the facts of the case, the injured party may bring a negligence or strict products liability claim [8]. Under negligence, the manufacturer of an AI-controlled system may be held liable if the system's components are not "reasonably competent" to perform together, thus breaching the manufacturer's duty of care to produce a non-defective product. However, under strict products liability (more on this later), a manufacturer may be held strictly liable if it places a defective product in the marketplace that causes injury to others. For such claims, as discussed throughout this paper, the system's input, output, and how the system is internally controlled are important aspects of the system to show whether the elements of negligence or strict products liability are relevant theories to pursue for disputes involving AI. For completeness, it is also necessary to briefly mention the theory of "malfunction liability". For a malfunction liability claim, direct evidence of the defective condition of a product or of the precise nature of the product's defect is not necessary. Here, the injured party can prove the existence of the defect through circumstantial evidence based on the occurrence of the malfunction or through evidence that eliminates any abnormal use of the product as well as reasonable secondary causes of the accident (on this point see Pagallo, U., *The Laws of Robots, CRIMES, Contracts and Torts*, Springer Press, 1165-117, 2013). This approach could be applicable, for example, if an AI-controlled system was damaged to the point where it is no longer possible to determine whether the AI system was designed with a defect based on having access to the actual system.

As another example of a systems and control theory approach to AI and law, the criminal justice system uses algorithms to determine whether a defendant convicted of a crime will likely be a repeat offender. A case in point is *Loomis v. Wisconsin* [9], in which the defendant challenged the State of Wisconsin's use of closed-source software (that was protected as a trade secret) in his sentencing [9]. The defendant alleged that using the software violated his constitutional right to due process because it prevented him from challenging the validity and accuracy of the risk assessment algorithm [9]. In this example, lack of access to the algorithm and lack of knowledge about how the algorithm functioned (that is, within the control structure of the system) may have violated the individual's constitutional, or more broadly, human rights to due process in a court proceeding.

Additionally, banks and credit unions decide who receives credit using a closed-loop feedback system that compares the desired outcome of the system with the initial input parameters (which are guided by the bank's lending policies and security regulations). By comparing the system's output to the system's input, the system may be adjusted to account for unplanned system performance, and in general, to adjust for any "noise" that may affect the performance of the system. The lack of system adjustments can lead to unexpected outcomes that may result in damages and thus disputes within law. This could occur, for example, with robots or autonomously operated vehicles if they lacked sufficient feedback on their actions in real-time, causing them to deviate from the desired system performance. Any "noise" affecting the system performance, that is, events not accounted for in the design of the system (but that were foreseeable), could be used by the plaintiff in a dispute involving an AI system to show, among others, that the system designer(s) was negligent.

System adjustments typically occur through the use of a closed-loop feedback system as shown in Figure 1, which is presented to summarize the discussion to this point. The Figure shows the basic control theory model in the context of data input to an AI system (represented by the black-box) and the resulting real-world behavior of the system. Based on the above discussion, we can generalize and say that there are three main areas of a systems and control theory approach to AI that should be of interest to legal scholars and to those with an interest to regulate AI: (1) the systems output, which could lead to actions that implicate the law, in which case not only would the systems output be relevant for resolving legal disputes, but (2) so too would the data, which serves as input to the system and how it is used within the "algorithmic structure" of the system be necessary in resolving a dispute, especially if the data led to biased decision making or physical damages, and (3) so too would the type, structure, and functioning of algorithm(s) operating within the black-box representation of the AI system provide important information for determining how the system was controlled and thus for assigning liability for damages.

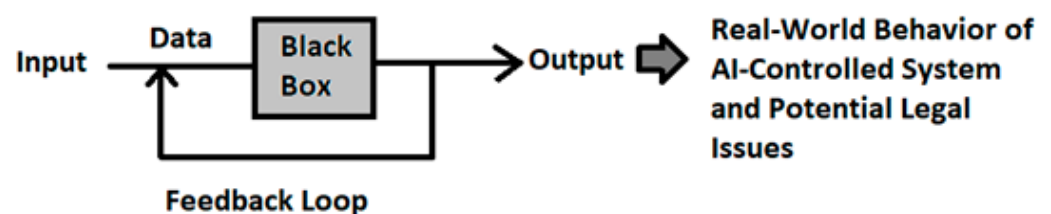


Figure 1. Example of a closed-loop system with an input in the form of data and an output represented by the real-world performance of the system.

Before going forward in the discussion, I should note that some aspects of a system may be less amenable to a control theory perspective than others. In particular, with reference to developing a law of AI, system components that lack algorithms to direct the actions of a system, or a system that lacks an identifiable control structure, will be more difficult to describe in the language of control theory than those systems that do contain such features. However, as AI becomes more widespread throughout society, industry,

and government agencies, even systems that lack a discernable physical structure (e.g., a system used to determine a person's credit rating or their eligibility for government-assisted housing) are using algorithms embedded within a computational infrastructure to achieve the systems goals. Thus, it appears that knowledge of systems and control theory will be applicable to a wide range of physical and nonphysical systems controlled by AI and therefore should be of interest to legal scholars and those tasked with regulating AI.

Summarizing the above discussion, in Figure 1, the "black-box" could represent any physical entity controlled by AI such as a robotic surgeon or autonomous vehicle [10] or the black-box could represent any non-physical entity such as a government agency or social organization [11], both of which take input in the form of data and then using algorithms make decisions that may affect an individual or a groups' fundamental rights under the law. In either case, both examples represent systems consisting of sub-components whose performance can be mediated through feedback loops; this results in a "controlled" system operating with little or no direct human input or supervisory control. For a physically controlled entity such as a robot, the actions of the robot as described by a systems and control theory approach could lead to harm to a human or damage to property, which represent external consequences of the systems output. On this point, consider a robot operating in a manufacturing facility. From a control theory perspective, a robot typically corrects its performance by comparing its output (for example, the position of a robot gripper or the amount of force it applies) with the desired end state of the system as indicated by its input parameters. If too much force is supplied by the robot's end effectors, which then leads to a negative outcome, when assigning liability for the robot's actions, the following could be considered: (1) the input parameters fed to the robot, (2) the algorithms used to adjust and control the motion of the robot and the amount of force it applied, (3) the mechanical structure of the robot, and (4) other systems and control-theory considerations discussed below.

A point to make is that an AI systems input–output control structure, while designed to operate efficiently within a specific environment, may not be able to account for all possible events that may affect system performance (this relates to the concept of foreseeability under negligence law). As an early example of this concept, in 1979, an American factory worker was killed by an industrial robot while working at an automotive plant. His family sued the manufacturers of the robot, alleging "... [negligence] in designing, manufacturing and supplying the storage system and in failing to warn [system operators] of foreseeable dangers in working within the storage area" [12]. Based on a jury decision, the court awarded his estate a sizeable financial award. To generalize, much of the performance of AI enabled technology (such as a robot) that is involved in a dispute could be described using concepts from systems and control theory; for example, the algorithms controlling a robot's motion (through feedback loops) could be used to show that the robot was either properly or improperly designed to work in the facility it was placed in or for the task it was required to perform. If it could be determined by examining an AI system's components and how they were controlled that the system (e.g., a robot) was improperly designed then in common law jurisdictions, and given specific case facts, a negligence or strict products liability claim could be pursued.

Additionally, to emphasize a former point, there may be circumstances when the doctrine of strict product liability may apply to a dispute involving AI, which begs the question of whether an understanding of systems and control theory is necessary for adjudicating the dispute. Unlike negligence, which requires the following elements—duty, breach of duty, and causation—the rules of strict products liability state that an individual will be strictly liable for harm caused to an individual without having to show fault or intent. For example, under strict products liability law, a manufacturer or distributor of a defective product (such as a da Vinci robotic surgeon) will owe an injured person compensation even if the defendant took reasonable steps to prevent the defect. Strict products liability claims require the establishment of one of three types of defects, two of which are discussed here, and both of which can occur with physically embodied AI. First,

even if an AI entity is manufactured according to its design specifications, if the design was considered dangerous, the use of the AI entity could lead to personal injuries and thus a strict products liability claim. Second, there could be a defect in the AI system that resulted from the manufacturing process that altered the AI product from its original design. For a strict products liability claim to be successful, it is necessary to establish that the injured party used the item as the manufacturer intended, the item contained a defect, and the item caused specific damages. As discussed throughout this paper, knowledge of systems and control theory provides a conceptual framework to guide the determination of whether an AI system was operating as expected or deviated from its desired performance. Further, systems and control theory can be used to show whether the product as operating in the world represented a dangerous design or, alternatively, whether harm to an individual resulted from a defect that occurred during the manufacturing process of the AI entity. Thus, even if a plaintiff's case is based on a theory of strict products liability, I argue that ideas from systems and control theory can still be used to show whether the elements of strict products liability apply.

2. The Black-Box and Closed-Loop Systems

To reiterate a main point discussed in this paper, basic knowledge of systems and control theory can assist the legal community in their understanding of the actions taken within the black-box representing an AI-controlled system, and such knowledge is useful in framing and resolving disputes. Such systems are often said to lack transparency due to the complexity of the system and of the algorithm(s) controlling the system [13]. On that point, one purpose of this paper is to try to demystify the control structure of a system in order to provide the legal community a new set of conceptual tools in which to understand how AI impacts the law. A closed-loop control system, also known as a *feedback control system*, uses the concept of an open-loop as its forward path but also has one or more feedback loops or paths between its output and its input [1]. The reference to "feedback" simply means that some portion of the output is returned "back" to the input to form part of the systems control [1,2]. Closed-loop systems are designed to automatically achieve and maintain the desired output condition by comparing it to the actual condition of the system. They do this by generating an error signal, which is the difference between the output and the reference input. In other words, a "closed-loop system" is a fully automatic control system in which its control action is dependent on the output in some way [2]. This aspect of a system, that is, that it operates automatically via a feedback control structure, is one of the features of a controlled system that make it difficult to assign liability to a human operator if damages result from the performance of the system.

For a dispute involving a system controlled by AI, if it is established that there was no manufacturing or design defect (or failure to warn of inherent dangers in a product), then the doctrine of strict products liability does not apply to the given dispute. However, from a control theory perspective, there are still two main ways in which an AI-controlled system may lead to undesirable performance that may implicate the law. First, the system may be open-loop and thus lack feedback control. In this case, the systems output is not compared to its input, and any resulting harm based on the performance of the system could be due to the lack of feedback control within the system (this implicates negligence in the design of the system and not strict products liability law). Second, the system may be closed-loop but still use faulty, corrupt, or misleading data as input (note that an open-loop system may also use similar corrupt data) or, within the system's black-box, use algorithms whose output is biased, thus leading to decisions that violate an individual's or group's rights. Either could have been the case with the computer system MiDAS, which was used to access a file in the Michigan Unemployment Insurance Agency database and then concluded incorrectly, and without human input, that an individual had defrauded the unemployment system and owed the state of Michigan restitution, penalties, and interest [14,15]. The system, still automated and without human involvement, collected the individual's assets by electronically intercepting the individual's tax refunds for a

two-year period. This is a clear example of a system that deviated from its desired system performance, which then resulted in harm to an individual.

If the MiDAS system lacked a feedback control loop between its input and output behavior, we can describe the system as being “open-loop”. Thus, in the *MiDAS* example, if the system was operating open loop, it would have lacked a feedback mechanism to compare the output decision (that insurance fraud occurred) with the input data contained within the state of Michigan’s database. With an open-loop system, because there is a lack of feedback between the system’s output and input, such lack of feedback may have resulted in inappropriate system performance. Figure 2 shows a highly simplified drawing of the functioning of MiDAS “as if” operating as an open-loop system.

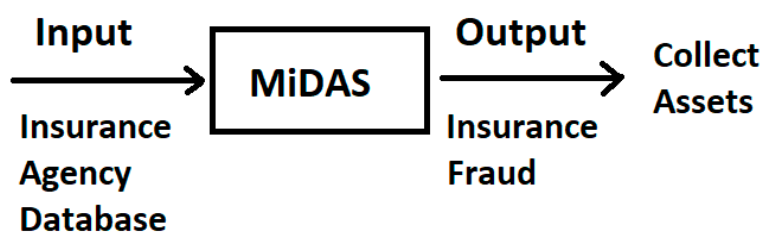


Figure 2. Example of an open-loop system lacking a feedback loop between the decision to collect assets and the algorithmic analysis of whether assets should be collected.

Without further detail on how MiDAS was designed to function in the government agency, it is possible that the system included a feedback-loop, allowing its output to be compared to its input values and then adjusted, in which case an analysis by a court would likely focus on the structure of the database, the algorithms determining that fraud occurred, the control structure of the system, and other system components comprising the AI’s “black-box”, which are discussed below. However, in either case (open- or closed-loop system), the above outcome in *Midas* can be described as a consequence of using a “black-box” AI-based system that may not be transparent, that may be complex (and beyond human understanding) [13], and that was implemented in a system that lacked proper human supervisory control. Not surprisingly, many AI techniques using algorithms are implemented using a black-box format, and often the assumption is that algorithms are objective and thus perform without biases. However, the input data may be “messy” real-world data or taken from biased sources; further, as indicated above, sources of “noise” may affect system performance [2]. That said, by framing a legal dispute involving AI in the language of systems and control theory, using the concept of open- or closed-loop systems, sources of error affecting system performance necessary to pursue a tort or other legal cause of action can be identified.

3. Sensors and Error in the Feedback Loop and System Control

The next section of the paper introduces additional topics necessary for an understanding of systems and control theory and that I argue are important concepts for developing a conceptual framework for how the law applies to AI. First, we note that AI-controlled systems equipped with feedback loops have sensors (for example, a potentiometer) that help monitor the state of a system in order to compare it with (or subtract it from) an input reference value. Typically, an error signal (see Figure 3), which is the required performance subtracted from the actual performance (done in a comparator), is amplified by a controller, which is the component of a system that monitors certain input variables and adjusts other output variables (or actuators) to achieve the desired system operation [1,2,16]. For example, consider the motion of a robot’s end effectors if the robot senses that its end effectors are deviating from the desired path (i.e., detects error); here, the controller may compensate by signaling movement corrections.

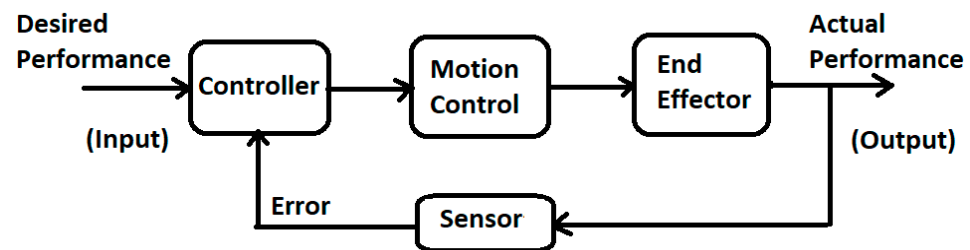


Figure 3. Example of closed-loop feedback system for a robot with an end effector.

From the discussion presented in the preceding section, another consideration for those interested in the regulation of AI is how the controller affects the system's performance. More specifically, this aspect of a system's structure should be of interest to the legal community in resolving disputes involving AI-controlled systems and for developing a conceptual framework for how the law applies to AI. Continuing the above "robot example", if the robot's end effectors are in close proximity to its desired end position, the controller may reduce the speed of the robot's arms or stop the movement altogether, so as to initiate a new set of instructions. Therefore, here, the closed-loop configuration of the system is characterized by a feedback signal, which contains a sensor, and using a controller adjusts the systems behavior. From this discussion, we can conclude that sensors and a controller are important concepts for law and AI, and particularly for algorithmically controlled systems, as how they function to control a system could lead to negative outcomes that implicate the law. Additionally, as a point to emphasize, knowledge of the control structure of a system will help courts determine which legal theories apply to a particular dispute involving AI.

For the above robot example, the magnitude and polarity of the resulting error signal (concerning the position of the robot's effectors) is directly related to the difference between the required position of the end effector and the position of the actual end effector [17]. Because a closed-loop system has some knowledge of the output condition (via a sensor), it is equipped to account for and control any system disturbances or changes in the conditions that may reduce its ability to complete a desired task [1]. Therefore, in the diagram below, I add a sensor and controller to the basic system and control theory model presented above. I note again that these system components act to control the behavior of a system and thus add to an understanding of how the law applies to AI, and particularly, algorithmically based systems.

As we can see in Figure 3, in a closed-loop system the error signal, which is the difference between the input signal and the feedback signal is fed to a controller to reduce system error and bring the output of the system back to a desired state. The accuracy of the output (which has relevance for legal disputes) depends, among other things, on the feedback path, which in general can be made very accurate using electronic control systems and circuits [2]. Additionally, as a basic idea, if the gain of the controller is too sensitive to changes in its input commands or signals, the system can become unstable and start to oscillate as the controller tries to over-correct itself [1]. This could lead to a breakage somewhere in the system being controlled. In general, gain is a proportional value that shows the relationship between the magnitude of the input and the magnitude of the output signal at a steady state [2]. Many systems contain a method by which the gain can be altered, for example, by providing more or less "power" to the system. Thus, as with a controller, the system's gain affects the performance of a system and should therefore be of interest to legal scholars formulating a law of AI and to courts resolving disputes involving algorithmically based AI systems.

4. Types of System Control

In control theory, there are two basic types of system control, both of which are important for understanding how the law applies to AI-controlled systems. These are

feedforward and feedback control [1,2]. With a feedback control system, the input to a feedback controller is the same as what the system is trying to control, and the controlled variable is “fed back” into the controller [1]. The cruise control in an automobile is an example of a feedback controller (malfunctioning cruise control systems have resulted in numerous legal disputes, see for example *Watson v. Ford Motor Company*, 389 S.C. 434 (2010)). Here, the controller measures the controlled variable, in this case the speed of the automobile, and by adjusting the engine throttle, the speed of the automobile (the systems output) is adjusted [18]. Among other things, the control of the system is provided through feedback loops operating within the system. In many systems, feedback control may result in intermediate periods where the controlled variable is not at the desired setpoint. With the cruise control example, if the automobile is on a downward slope, the automobile would be instructed to slow down by adjusting its throttle. If the automobile’s speed falls below the desired speed, the controller will act to adjust the throttle to increase the speed. We can see from this simplified example that a controller, as part of the black-box representation of a system, is essential for regulating a system’s performance. Therefore, the design of a controller and how it performs under various conditions should be of interest in a dispute involving an AI system that is operating closed-loop. For example, if a controller malfunctions, which leads to damages, among other things, a strict products liability or negligence claim may be pursued depending on the facts of the case. Additionally, if the controller functions as designed but damages still occur, then the court may look to other aspects of the systems design to determine the reason for the accident and if the manufacturer was liable under various legal theories [19] or, alternatively, if there was operator error.

With feedforward control, disturbances are measured and accounted for before they have time to affect the system [1]. In the cruise control example, using GPS a feedforward system may detect that the automobile is approaching the downslope of a hill and will automatically adjust the automobile’s throttle before it exceeds its desired speed. However, the difficulty with feedforward control is that the effect of the disturbances on the system must be accurately predicted, which may be difficult and there must not be any surprise disturbances. In a legal dispute involving AI, identifying surprise disturbances that could lead to negative outcomes, and which any human in the design of the system should have predicted beforehand would occur, would be of interest to courts in assigning liability. In the context of a court action, one element of negligence in common law jurisdictions is proximate cause. Here, the defendant is responsible for those harms the defendant could have foreseen would result through his or her actions [20]. If the system was not designed to account for system disturbances, a question under the law of negligence is whether the defendant should have foreseen that the control aspect of the system was inadequate to account for such disturbances.

To achieve the benefits of feedback control (controlling unknown disturbances and not having to know exactly how a system will respond to disturbances) and the benefits of feedforward control (responding to disturbances before they can affect the system), there are combinations of feedback and feedforward control that can be used [1]. Some examples of where feedback and feedforward control can be used together are dead-time compensation and inverse response compensation [2]. Dead time is the delay from when a controlled output signal is issued until when the measured process variable first begins to respond [1,2]. Inverse response compensation involves controlling systems where a change at first affects the measured variable one way but later affects it in the opposite way [1,2]. As can be imagined, it is difficult to control this system with feedback alone; therefore, a predictive feedforward element is necessary to predict the reverse effect that a change will have in the future.

For feedback controllers, there are a few types to note in this introductory tutorial on control theory and law. As an example of a simple controller, consider a thermostat that just turns the heat on if the temperature falls below a certain value and off it exceeds a certain value (on-off control). Generally, the binary nature of the system control allows

system errors to be easily identified in resolving disputes. Another type of controller is a proportional controller [21]. With this type of controller, the controller output (control action) is proportional to the error in the measured variable. As noted above, the error is defined as the difference between the current value (measured) and the desired value (setpoint). If the error is large, then the control action is large. As a summary of the ideas introduced above, let e = error, and K represent gain, then . . .

$$c(t) = K_c * e(t) + c_s$$

where . . .

$e(t)$ represents the error in the system control,

K_c represents the controller's gain, and

c_s represents the steady state control action necessary to maintain a variable at the steady state when there is no error.

The gain K_c will be positive if an increase in the input variable requires an increase in the output variable (direct-acting control), and it will be negative if an increase in the input variable requires a decrease in the output variable (reverse-acting control) [2]. Although the functioning of proportional control should not be difficult to explain in a court proceeding as it is fairly straight-forward, it has technical drawbacks. The largest problem is that for most systems, it will never entirely remove error. This is because when error is "0", the controller only provides the steady state control action so the system will settle back to the original steady state (which is probably not the new set point that the system should be at). To make the system operate near the new steady state, the controller gain, K_c , must be very large so the controller will produce the required output when only a very small error is present. Having large gains can lead to system instability or can require physical impossibilities like infinitely large valves for a mechanical system [1,2]. Admittedly, a discussion of alternates to proportional control, which are proportional-integral (PI) control and proportional-integral-derivative (PID) control, is beyond the scope of this paper. However, I should note in this tutorial that PID control is commonly used to implement closed-loop control; therefore, an understanding of PID control would be beneficial to legal scholars and legislators interested in regulating law and AI [21]. Another important concept for law and AI-controlled systems is the transfer function of any electrical or electronic control system, which is the mathematical relationship between the systems input and its output that describes the behavior of the system [1,2]; thus, it should be of interest to legal scholars with an interest in law and AI.

5. Example of a Case

The next section of the paper focuses on a brief analysis of a case that involved a system operating with several levels of feedback control implemented through the use of algorithms and mechanical parts serving as subcomponents within the system. In *Jarvis v. Ford Motor Company*, a Ford Aerostar driven by Jarvis led to an accident that injured Jarvis when the Aerostar suddenly accelerated [22]. Jarvis argued that the Aerostar accelerated even though she had not depressed the accelerator and that pumping the breaks did not stop the van. Among other legal theories, Jarvis claimed that Ford was negligent and should be held strictly liable for the design of the Aerostar's cruise control mechanism [22]. Here, I should note that the cruise control mechanism is an example of a feedback control system. According to the court, to prove negligence, Jarvis was not required to establish what specific defect caused the Aerostar to malfunction, and circumstantial evidence concerning the functioning of the system itself could be dispositive [22]. Regardless of whether knowledge of a specific defect was known or whether circumstantial evidence was available, an understanding of how the system operated and how it was controlled through feedback loops and other control variables was necessary for resolving the dispute and assigning liability.

From a control theory perspective, an issue of interest for the court was the functioning of the automobile's braking and cruise control system and whether they either malfunctioned or were defectively designed, thus implicating strict products liability law. Describing the control system structure of the braking system, an expert witness for Jarvis testified that the Aerostar had vacuum power brakes that worked by drawing their vacuum from the engine but that the engine did not create the necessary vacuum when accelerating full throttle [22]. From a systems perspective, within the braking system, a check valve is used to trap a reservoir of vacuum for use when the engine vacuum is low. However, in *Jarvis*, an expert witness testified that the reserve could be depleted after one-and-a-half hard brake applications, which can occur if an automobile suddenly accelerates and in response the driver reacts by pumping the brakes [22]. As a further description of the control structure of the system, a check valve is designed to draw out air that is trapped in the brake booster without letting additional air enter the brake cylinder [22]. More specifically, the check valve connects the body of the brake booster (which increases the force applied from the brake pedal to the master cylinder) to a vacuum hose and is a safety solution that should allow the brakes to work, even if the engine is shut off [22].

Another expert witness for Jarvis offered a different theory on how the accident occurred, again, describing the accident in the language of systems and control theory. The expert witness argued that unintended electrical connections can cause current to run to the cruise control "servo", opening the throttle and making the vehicle accelerate [22]. Describing the "control structure" of the cruise control system of the automobile, the cruise control servo opens the automobile's throttle by means of a vacuum mechanism that has two valves, the vacuum, described as the "vac" valve, and the vent valve, which are controlled by wires attached to the "speed amplifier" [22]. The expert witness testified that if there was a simultaneous short in the vent and the vac wires, the cruise control servo would open the throttle without the driver pressing the accelerator [22]. The dispute among the experts testifying in court was whether these kinds of electrical malfunctions could spontaneously occur and, if so, whether a failsafe mechanical device called a "dump valve" would, despite these malfunctions, disengage the cruise control servo as soon as Jarvis pressed the brake pedal, bringing the sudden acceleration to an end and the vehicle to a stop [22].

Given the above two explanations offered by expert witnesses for Jarvis, both focusing on how the accident may have occurred, and both using a system and control theory perspective, the court had to connect the design and functioning of the system to the elements of the legal theories pursued by Jarvis. One legal theory offered by Jarvis was negligent design [22]. The instructions to the jury on negligent design comprised three elements: (1) that the cruise control system in the 1991 Aerostar was defective when put on the market by Ford, (2) that the defect made it reasonably certain that the vehicle would be dangerous when put to normal use, and (3) that Ford failed to use reasonable care in designing the cruise control system or in inspecting it or testing it for defects. Further, even if Ford used reasonable care in designing, inspecting, and testing the cruise control system in the Aerostar, Ford learned of the defect before putting the product on the market and failed to correct it [22]. The court reasoned that though the occurrence of the accident was not proof of a defective condition, a defect may be inferred from proof that the product did not perform as intended by the manufacturer [22]. Interestingly, the court testimony by experts focused on how the cruise control and braking system operated within the control structure of the automobile, thus indicating the importance of systems and control theory in framing the dispute.

6. Conclusions and Overview of a Systems and Control Theory Approach for Law and Artificial Intelligence

In summary, this paper discussed systems and control theory and their application to AI-controlled systems. In that context, Figure 4 shows the control-theory framework presented in this paper with selected areas of law that relate to different aspects of the systems control structure identified. For example, the input data to a system is not only

used by algorithms within the system to produce an output but the input may be protected by patent, copyright, and trademark law [23]. Further, the algorithms functioning within the “black box” structure of an AI-controlled system may result in decisions that could impact an individual’s constitutional or human rights (e.g., consider the case of algorithmic bias that may impact the rights of a protected group). Additionally, in some cases, algorithms may be patent-protected subject matter or protected as a trade secret. Additionally, copyright law might protect the written form of an algorithm or possibly the structure of a database used for training an algorithm operating within an AI system. Further, the system’s output, which is often the focus of a legal dispute, can result in actions that challenge civil and criminal law, constitutional law, and human rights law, and may itself be protected by intellectual property law [23]. And as discussed earlier in this paper, negligence and the doctrine of strict products liability may apply to disputes involving AI-controlled systems given the specific facts of the case. Other rights may also be impacted when using an AI-controlled system; for example, if the data used by the system in decision making is of a personal nature, data protection laws may be triggered. Additionally, other legal constraints may come into play in some jurisdictions if data are owned by a public administration or government agency in which case they may be protected as a state secret [24].

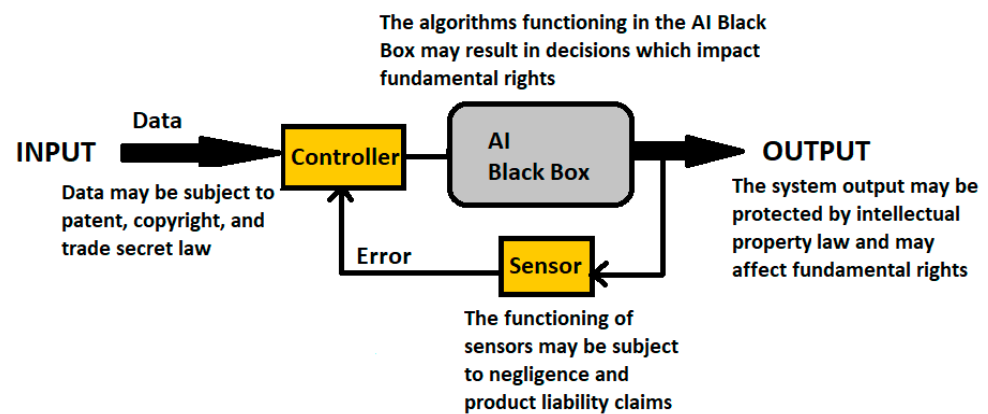


Figure 4. A systems and control theory model for law and artificial intelligence with areas of law that relate to different system components identified.

Considering Figure 4, its purpose is to highlight the idea that the law applies to different aspects of systems controlled by AI. However, currently, the technology and control structure operating between the input and output of a system are often not the focus of the legal community in discussions of how the law applies to AI or of courts when resolving disputes involving AI systems (with the exception of tort actions involving braking and cruise control malfunctions discussed above). However, as AI-controlled systems become more prominent, more litigation can be expected that will require an understanding of system components and how they are controlled. This paper is a step in the direction of elucidating the concepts of systems and control theory by presenting a tutorial to aid the legal community in resolving disputes involving AI and in developing a conceptual framework for law as applied to AI.

It should be noted that the examples presented in this paper described AI-controlled systems making decisions that do not involve moral or ethical considerations, yet AI is being used in contexts where such decisions may be necessary. This means that the control structure of the system must be designed for such contingencies. While a full discussion of this topic is beyond the scope of this introduction to control theory and law, a few preliminary points can be made. One solution to address moral dilemmas that AI systems may encounter would be to design the AI system to alert a human of the moral dilemma and then allow the human to decide the remedial course of action. However, under some circumstances, there may not be sufficient time for this solution to be practical,

in which case, I would argue that the control structure of an AI system could respond with sufficient speed to ensure that the least harm from a set of options would be implemented. Here, public policy on ordering the various contingencies based on agreed upon policy goals would need to be determined a priori and programmed or hardwired into system components, in this way influencing the control structure of the system and its output. Additionally, standards, such as shown in Article 53 of the Act of the European Commission (which is on the regulation of AI), need to be considered for AI systems that may make moral or ethical decisions.

To reiterate the main point of the paper, given that AI-controlled systems are often designed with several feedback loops in their control structure and that they have sensors and controllers that direct their performance, concepts from systems and control theory provide a useful way to conceptualize legal disputes involving AI-controlled systems and particularly those that are algorithmically based and that are operating autonomously. While the control structure of AI systems may be complex in some cases, the fundamental ideas as discussed in this paper are still relevant for understanding the performance of the system, and whether its design and how the system functioned were a factor in any harm resulting from the system's performance. Further, recent papers discussing law and AI have emphasized the lack of transparency for "black-box" AI-controlled systems [13]. As argued here, systems and control theory provides the legal community a look into what is inside the black box, which should provide value to the discussion of how the law can be used to regulate AI.

There are of course, limitations to the current paper. The examples here focused primarily on systems with one feedback loop, yet most systems controlled by AI consist of numerous subsystems with multiple control loops operating within the structure of the system. However, even for more complex systems than those discussed in this paper, the basic ideas presented in this paper still apply and are useful for understanding how the law relates to disputes involving AI. Further, in terms of practical guidelines, it is often stated that transparency is important for AI systems, so a control theory approach to AI can be used to identify several areas of the system where an explanation of the systems performance and control structure would address the issue of transparency. For example, knowledge of how the controller operates, the nature of feedback loops, and the transfer function describing how the system is controlled would help address the issue of transparency in algorithmic decision making.

Future directions of the approach described in this paper are severalfold: (1) it is necessary to more carefully map systems and control theory concepts to the elements of legal theories used to litigate disputes; (2) as AI proliferates into society, it is necessary to develop law that takes into account how an AI system is actually controlled, which is to say, the control structure of a system and how that leads to decisions that may challenge the law; and (3) it is necessary to inform those with an interest in law and technology how systems and control theory provide critical concepts in understanding how AI systems may challenge the law. As one final thought, Lawrence Lessig wrote eloquently that code is law [25]. From the analysis presented in this paper, the law as applied to AI should be shaped by knowledge of controllers, transfer functions, the systems gain, feedback loops, and other control system concepts. That is, just as the law regulates human activities, and code regulates cyberspace, the computational and mechanical architecture of an AI system and its control structure informs the system as to what it can do and thus regulates its activities. We must take that knowledge into account going forward as we develop a framework for understanding how the law applies to AI-controlled systems.

Funding: This research received no external funding.

Acknowledgments: I thank the reviewers for providing valuable comments to an earlier draft of the paper and Ugo Pagallo, co-editor of the special issue, for his timely comments and thoughtful suggestions.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Nise, N.S. *Control Systems Engineering*, 7th ed.; Wiley Publisher: Hoboken, NJ, USA, 2015.
2. Kossiakoff, A.; Biemer, S.M.; Seymour, S.J.; Flanigan, A. Wiley Series in Systems Engineering and Management. In *Systems Engineering Principles and Practice*, 3rd ed.; Wiley Publisher: Hoboken, NJ, USA, 2020.
3. Yu, R.; Ali, G.S. What's Inside the Black Box? AI Challenges for Lawyers and Researchers. *Leg. Inf. Manag.* **2019**, *19*, 2–14. [[CrossRef](#)]
4. Von Bertalanffy, L. An Outline of General System Theory. *Br. J. Philos. Sci.* **1950**, *1*, 134–165. [[CrossRef](#)]
5. Von Bertalanffy, L. The History and Status of General Systems Theory. *Acad. Manag. J.* **1972**, *15*, 407–426.
6. Weber, M.; Kowalkiewicz, M.; Weking, J.; Bohm, M.; Krcmar, H. When Algorithms Go Shopping: Analyzing Business Models for Highly Autonomous Consumer Buying Agents. In *Entwicklungen, Chancen und Herausforderungen der Digitalisierung: Band 1, Proceedings der 15. Internationalen Tagung Wirtschaftsinformatik*; Gronau, N., Krasnova, H., Pousttchi, K., Heine, M., Eds.; Institut für Wirtschaftsinformatik und Digitale Gesellschaft e.V.: Potsdam, Germany, 2020; pp. 1–16.
7. Bafghi, E.P. Clustering of Customers Based on Shopping Behavior and Employing Genetic Algorithms. *Eng. Technol. Appl. Sci. Res.* **2017**, *7*, 1420–1424. [[CrossRef](#)]
8. Barfield, W. Liability for Autonomous and Artificially Intelligent Robots. *Paladyn J. Behav. Robot.* **2018**, *9*, 193–203. [[CrossRef](#)]
9. *Loomis v. Wisconsin*, 881 N.W.2d 749 (Wis. 2016), cert. denied, 137 S.Ct. 2290 (2017).
10. Dekoulis, G. (Ed.) *Robotics, Legal, Ethical, and Socioeconomic Impacts*; InTech Publisher: London, UK, 2017.
11. Picazo-Vela, S.; Fernandez-Haddad, M.; Luna-Reyes, L.F. Opening the Black Box: Developing Strategies to Use Social-Media in Government. *Gov. Inf. Q.* **2016**, *13*, 693–704. [[CrossRef](#)]
12. Young, B. The First “Killer Robot” Was Around Back in 1979. Available online: <https://science.howstuffworks.com/first-killer-robot-was-around-back-in-1979.htm> (accessed on 22 August 2021).
13. Larsson, S.; Heintz, F. Transparency in Artificial Intelligence. *Internet Policy Rev.* **2020**, *9*, 1–16. [[CrossRef](#)]
14. Yang, B.; Gilles, H.; Bin, K. Artificial Intelligence and Fraud Detection. In *Innovative Technology at the Interface of Finance and Operations*; Babich, V., Birge, J., Hilary, G., Eds.; Springer Series in Supply Chain Management; Springer Nature: Berlin/Heidelberg, Germany, 2020; Available online: <https://ssrn.com/abstract=3738618> (accessed on 28 July 2021).
15. *Bauserman v. Unemployment Insurance Agency*, 931 N.W.2d 539 (2019), 503 Mich. 169.
16. Bryan, L.A.; Bryan, E.A. *Programmable Controllers: Theory and Implementation*, 2nd ed.; Industrial Text Co. : Atlanta, GA, USA, 1997.
17. Sciavicco, L.; Bruno Siciliano, B. *Modelling and Control of Robot Manipulators (Advanced Textbooks in Control and Signal Processing)*, 2nd ed.; CreateSpace Independent Publishing Platform: Livermore, CA, USA, 2001.
18. Rajamani, R. *Vehicle Dynamics and Control (Mechanical Engineering Series)*, 1st ed.; Springer: Berlin/Heidelberg, Germany, 2006.
19. *Wernsing v. General Motors Corporation*, 298 Md. 406 (1984), 470 A.2d 802.
20. Beck, S. Intelligent Agents and Criminal Law—Negligence, Diffusion of Liability and Electronic Personhood. *Robot. Auton. Syst.* **2016**, *85*, 138–143. [[CrossRef](#)]
21. Schei, T.S. A Method for Closed Loop Automatic Tuning of PID Controllers. *Automatica* **1992**, *8*, 587–591. [[CrossRef](#)]
22. *Jarvis v. Ford Motor Company*, 283 F.3d 33 (2002).
23. Barfield, W.; Pagallo, U. *Advanced Introduction to Law and Artificial Intelligence*; Edward Elgar Publisher: Cheltenham, UK, 2020.
24. An Overview of the Data Protection and Security Regimes of the Peoples Republic of China. Available online: <https://www.steptoe.com/images/content/6/2/v4/6266/China-Data-Protection-Regime-Overview-05112017-Final.pdf> (accessed on 20 August 2021).
25. Lessig, L. *Code and Other Laws of Cyberspace*; Basic Books: New York, NY, USA, 1999.