INTERNATIONAL LAW STUDIES

Published Since 1895

Lethal Autonomous Weapon Systems: Translating Geek Speak for Lawyers

Linell A. Letendre

96 INT'L L. STUD. 274 (2020)

Volume 96 2020

Published by the Stockton Center for International Law

ISSN 2375-2831

Lethal Autonomous Weapon Systems: Translating Geek Speak for Lawyers

Linell A. Letendre[*]

CONTENTS

* Dean of the Faculty, United States Air Force Academy and former Permanent Professor of the Department of Law, United States Air Force Academy.

The thoughts and opinions expressed are those of the author and not necessarily those of the U.S. government, the U.S. Department of the Navy, the U.S. Naval War College, the U.S. Department of the Air Force, or the U.S. Air Force Academy.

I. INTRODUCTION

> *An engineer, a commander, and a lawyer all met at the pearly gates of heaven. Their only entry requirement was that they be able to speak the same language regarding lethal autonomous weapon systems . . . .*
>
> *Query: How long did it take them to pass through the gates?*

While the above sounds like the beginning of a trite joke, the sad and unfortunate aspect is that all too often lawyers' eyes glaze over as engineers begin speaking about reliability, control theory, and open architecture design, while engineers tune-out at the mere mention of international law principles and go into a comatose state when the Latin phrases begin. However, with the growing utilization of robotics and autonomous systems on the battlefield, lawyers and engineers must learn to speak the same language—or at the very least, understand the basics of the other's world.

Why the imperative? Why can engineers not just keep designing systems based on operator requirements and military lawyers not just keep evaluating the legality of the weapons systems once the systems are built? The simple answer is that autonomy is different. Advances in autonomy have the potential to move the human warrior further and further out of the control loop and to leave more and more decisions up to the machine. The catch is that the laws of war are to be assessed and implemented by people, not machines. Translating the legal requirements for the use of force cannot be an afterthought in the development of autonomous weapon systems; it must be built-in from the beginning. The better engineers understand international law requirements and lawyers understand the technical limitations of autonomous systems, the more likely we are to develop lethal autonomous weapon systems that comply with the law.

To further such understanding, this article demystifies robotics and autonomy for lawyers. While a lawyer will not be qualified to build a weapons system after reading this article, he or she should at least be able to follow a conversation surrounding robotics and ask intelligent questions of robotics engineers and the operators. The article starts with the basics by defining robots, autonomy, and operator (warfighter) terminology for unmanned platforms. Then, it explores how robotic and autonomous systems work by explaining control systems and robotic learning and reasoning. After most sections, the article provides a general overview of the topic discussed and explains why it is important for lawyers to understand the key issues related to this topic. As such, there are a series of questions to aid lawyers in asking relevant and meaningful questions of robotic engineers.

Returning to the opening question, were there not two other people besides the lawyer at the pearly gates? Yes, the engineer and the commander joined the lawyer at the heavenly gates. A companion article, *Lethal Autonomous Weapons: Translating Legal Jargon for Engineers*,[1] provides an overview of the international law governing the use of autonomous weapon systems in combat. Similar to this article, *Translating Legal Jargon for Engineers* provides questions that an engineer may use to clarify design specifications with lawyers during the development of autonomous weapon systems.[2] The combination of these two articles provides a necessary foundation for lawyers and engineers to understand autonomous weapon systems.

The remaining person, the commander, is the individual with the most at stake in ensuring that this conversation between lawyers and engineers occurs. Even as human operators are pushed further from decision loops in autonomous systems, the commander will remain legally (and at times criminally) responsible for the proper employment of these weapon systems in combat.[3] Thus, commanders should strive to pull engineers and lawyers into the same room when developing requirements for future autonomous systems and force a dialogue between them. To encourage meaningful dialogue, commanders should review the suggested list of questions that a military lawyer could ask a robotics engineer and vice versa. If a commander hears these types of questions, he or she can have confidence that all the parties to the critical determination of whether an autonomous weapon system is legal (either per se or in its employment) are speaking the same language.

1. Linell A. Letendre, *Lethal Autonomous Weapon Systems: Translating Legal Jargon for Engineers*, 2016 INTERNATIONAL CONFERENCE ON UNMANNED AIRCRAFT SYSTEMS 795 (2016).

2. Because few engineers are regular readers of law reviews, the author published that article in the 2016 International Conference on Unmanned Aircraft Systems proceedings. *See supra* note 1.

3. For a discussion of criminal liability in the context of autonomous weapon systems, see Christopher M. Ford, *Autonomous Weapons and International Law*, 69 SOUTH CAROLINA LAW REVIEW 413, 463–70 (2017).

II. THE BASICS OF ROBOTICS AND AUTONOMY

A. What Is a Robot?

If you polled a dozen robotics engineers and asked for a definition of a robot, you would get at least a dozen answers. Two Czechoslovakian playwrights first coined the term robot in a 1921 play by combining the terms "rabota" (meaning obligatory work) and "robotnik" (meaning serf).[4] Since that time, literary and cinematic depictions have morphed the populace's view of robots from the helpful type like the beeping R2-D2 in *Star Wars* and the loveable-but-tough Baymax in *Big Hero 6* to the frightening Cyberdyne Systems Model 101 in *Terminator*. This distortion of robots provides little assistance when trying to reconcile what makes a robot a robot vice a simple machine.

At its essence, a robot is a "machine that senses, thinks, and acts" in the physical world.[5] First, a robot senses both the physical environment around it as well as its internal state.[6] A robot relies on exteroceptive sensors to make sense of its outside or external environment and proprioceptive sensors to monitor its internal status.[7] Armed with that information, a robot is then capable of making a decision based on those inputs using its controller (brain) or computer processing capabilities. Finally, a robot acts in some fashion to accomplish a goal. It can do this by physically manipulating its environment using effectors (also known as end-effectors) or by traveling (often referred to as locomotion by engineers) to a particular point or place. For lethal autonomous weapon systems, the ultimate action for a robot is the delivery of a combat effect on the adversary or its equipment. The U.S. Department of Defense (DoD) defines an autonomous weapon system as a "weapon system that, once activated, can select and engage targets without further intervention by a human operator"[8] This definition reflects the requirement of a combat effect.

4. MAJA J. MATARIĆ, THE ROBOTICS PRIMER 1 (2007).

5. GEORGE BEKEY, AUTONOMOUS ROBOTS: FROM BIOLOGICAL INSPIRATION TO IMPLEMENTATION AND CONTROL 2 (2005).

6. MATARIĆ, *supra* note 4, at 1–26.

7. BEKEY, *supra* note 5, at 11.

8. U.S. Department of Defense, Directive 3000.09, Autonomy in Weapon Systems 13 (2012, incorporating Change 1, May 8, 2017), https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf [hereinafter DoD Directive 3000.09].

B. Automation . . . Autonomy . . . Automaton . . . Oh My!

At minimum, military lawyers working with engineers and operators need to get their terms right. Let us start with a simple declaration: automation is *not* the same as autonomy. Master this distinction and no one will accuse you of being a robotic neophyte. Automation is simply using a machine to do a particular process, often one previously done by humans.[9] Autonomy, on the other hand, denotes a system capable of operating independently for some period and making its own decisions without direct human supervision.[10] Renowned American roboticist George Bekey described autonomous robots as "intelligent machines capable of performing tasks in the world *by themselves*."[11] Thus, the difference between a machine and a robot with some level of autonomy is that a machine's actions are completely controlled and designed by humans, whereas autonomous robots may take advice and direction from humans, but humans may not control every action or decision.[12] In short, unlike machines, autonomous systems can make an independent decision and then act upon it.[13]

Robotic systems will vary in their level of autonomy—from full human control to semi-autonomous to full-autonomy—based on the extent of human interaction between the machine and the robot. Further, the level of autonomy realistically changes depending on the particular subsystem, the type of mission, or simply which part of the mission is being performed.[14] In 2012, the Defense Science Board recommended abandoning the use of "levels of autonomy" altogether.[15] However, the Defense Department has not implemented this recommendation and the concept continues to appear in DoD studies and reports. Thus, lawyers working in this field should be familiar with several of the major approaches regarding levels of autonomy.

Early roboticists divided autonomous robots into three categories or types: scripted, supervised, and intelligent.[16] Scripted robots follow pre-

9. STAN GIBILSCO, CONCISE ENCYCLOPEDIA OF ROBOTICS 15 (2002).

10. BEKEY, *supra* note 5, at 1.

11. *Id.* at xiii (emphasis added).

12. MATARIĆ, *supra* note 4, at 3.

13. *Id.* at 26.

14. DEFENSE SCIENCE BOARD, U.S. DEPARTMENT OF DEFENSE, TASK FORCE REPORT: THE ROLE OF AUTONOMY IN DOD SYSTEMS 62 (2012), https://apps.dtic.mil/dtic/tr/fulltext/u2/a566864.pdf.

15. *Id.*

16. JOSEPH A. ANGELO JR., ROBOTICS: A REFERENCE GUIDE TO THE NEW TECHNOLOGY 252 (2007).

planned mission descriptions, such as guided missiles and fire-and-forget weapons. Supervised robots require a human to oversee various aspects of the mission and provide decision making at certain points. Intelligent robots maintain complete control of decision making without the need for human intervention. These categories have now given way to descriptions of the levels of autonomy from different viewpoints and in differing contexts. The approaches common within military systems are detailed below.

The Defense Department outlined four levels of autonomy in its 2011 *Unmanned Systems Integrated Roadmap*.[17] This approach focused on whether the machine or the human makes the decision. The 2011 document describes the four levels as (1) human operated (where the human makes all decisions), (2) human delegated (where the system can perform those functions allowed by the human), (3) human supervised (where the unmanned system conducts an array of activities under the supervision of a human operator, including initiating its own actions based on the machine's assessment of the environment), and (4) fully autonomous (where the system functions completely independently of a human operator other than the human setting the overarching goals of the mission). Figure 1, below, details these levels.

Level	Name	Description
1	Human Operated	A human operator makes all decisions. The system has no autonomous control of its environment although it may have information-only responses to sensed data.
2	Human Delegated	The vehicle can perform many functions independently of human control when delegated to do so. This level encompasses automatic controls, engine controls, and other low-level automation activated or deactivated by human input.
3	Human Supervised	The system can perform a wide variety of activities when given top-level permissions or direction by a human. Both the human and the system can initiate behaviors based on sensed data, but the system can do so only if within the scope of its currently directed tasks.
4	Fully Autonomous	The system receives goals from humans and translates them into tasks to be performed without human interaction. A human could still enter the loop in an emergency or change the goals, although in practice there may be significant time delays before human intervention occurs.

Figure 1: Levels of Autonomy from 2011 DoD Unmanned Systems Integrated Roadmap

17. U.S. DEPARTMENT OF DEFENSE, UNMANNED SYSTEMS INTEGRATED ROADMAP FY 2011–2036 (2011), tbl. 3, at 46, https://info.publicintelligence.net/DoD-UAS-2011-2036.pdf.

In 2012, the Defense Department published DoD Directive 3000.09, which somewhat narrowed this approach. This Directive primarily addressed the development and use of autonomous and semi-autonomous functions in weapon systems.[18] The Directive also defined an autonomous weapon system as "[a] weapon system that, once activated, can select and engage targets without further intervention by a human operator."[19] In contrast, semi-autonomous systems are "intended to only engage individual targets or specific target groups that have been selected by a human operator,"[20] whereas human-supervised autonomous weapon systems can select targets independently, but are "designed to provide human operators with the ability to intervene and terminate engagements."[21]

A second framework often seen in military robotics is the autonomy levels for unmanned systems (ALFUS) approach.[22] This model, displayed in Figure 2 below, uses a three-axis approach for assessing the level of autonomy for a system.

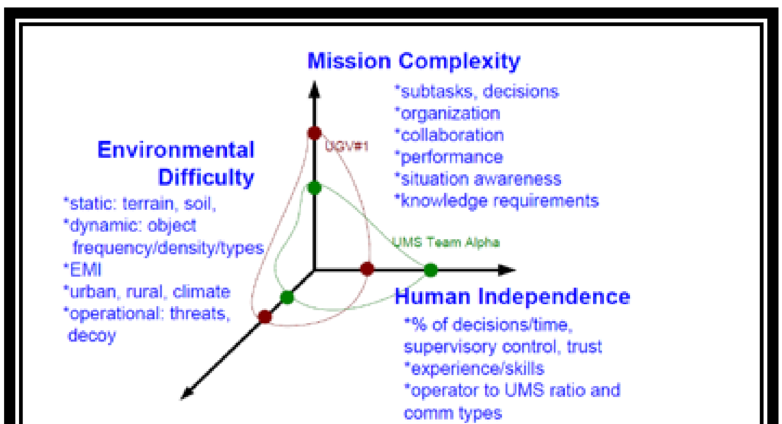

Figure 2: Autonomy Levels for Unmanned Systems

18. DoD Directive 3000.09, *supra* note 8, at 1.
19. *Id.* at 13.
20. *Id.* at 14.
21. *Id.*
22. HUI-MIN HUANG ET AL., A FRAMEWORK FOR AUTONOMY LEVELS FOR UNMANNED SYSTEMS (ALFUS) (2005), https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=822679.

This approach allows engineers to demonstrate that the level of autonomy for a particular system is a product of many factors, such as the complexity of the assigned missions (mission complexity), the difficulty of the environment in which the system will operate (environmental difficulty), and how much of the supervision and how many of the decisions will be made by the human operator vice the machine (human independence).[23] While the ALFUS model has been helpful in allowing developers to assess how levels of autonomy may change in different contexts, the model also has several drawbacks for the test and evaluation community. In particular, the complexity of the model and the need for developers to refine it continually, demonstrate some of the model's limitations.[24]

A third DoD framework eliminates the need to evaluate how autonomy changes in various missions or environments and instead focuses on the overall autonomy potential of the system. It is appropriately named the non-contextual autonomy potential (NCAP) model and ranges from an autonomy level (AL) of zero for non-autonomous systems to a maximum of three for fully autonomous systems. For example, teleoperated or unmanned systems operated by remote control are designated as an AL 0. Such systems simply collect data or information about the environment but do not process the information outside of potentially relaying it.[25] If the system creates a model with the data it collects, the NCAP model considers the system semi-autonomous.[26] At AL 2, the system is capable of planning based on its self-created model and is considered an autonomous system. To be "fully autonomous" and receive an AL 3 designation, the NCAP model requires that the system be able to plan and then act on that plan without human involvement.[27] Note that the NCAP model is very similar to the four levels (1-4) of autonomy specified in the DoD *Roadmap*, except that as opposed to looking at what the human is doing, NCAP focuses on what the machine has the potential to accomplish on its own.

---

23. *Id.* at 5.

24. PHILLIP J. DURST & WENDELL GRAY, LEVELS OF AUTONOMY AND AUTONOMOUS SYSTEMS PERFORMANCE ASSESSMENT FOR INTELLIGENT UNMANNED SYSTEMS 14 (2014), https://apps.dtic.mil/dtic/tr/fulltext/u2/a601656.pdf.

25. *Id.* at 17.

26. *Id.*

27. *Id. at* 17–18.

Questions lawyers should ask engineers about autonomy:

- How would you describe the range of levels of autonomy this weapon system possesses?
- Does the potential autonomy of the system vary depending on the environment complexity or the given mission? If so, how?
- How have scientists tested these levels of autonomy?
- Under a typical or expected concept of operations for this system, how will the autonomy level change? What is the maximum expected autonomy level?

C. Is it a Bird, a Plane, a UAV, or a RPA? A Guide to Robotic Alphabet Soup

Good military lawyers understand the importance of learning the specific language used by operators and the robotics and autonomous systems arena is no different. Consider walking into a conversation with a Marine UAV commander that sounds like this:

> Our current UAS in the USMC is a Group III RQ-7B. All UXS are a complex SoS comprised of the GCS and GDT, which connects the UAV and GCS through an RF datalink system, sometimes via Ku band. Group IV and V UAS operate BLOS using SATCOM and Group I and II utilize LoS datalinks, which can be L, C, Ku, Ka, or even traditional Wi-Fi. In terms of manpower, the UAS is typically manned by an AVO, who controls the AV via keyboard and mouse, the MPO, who controls the POP-300D sensor and LD, a UAC, who is an officer, and fully-trained RPA, and is assisted by an IS who synthesizes outside collection sources to tell the operators where to place the camera most effectively. Finally, we need the FSRs to perform MX.[28]

What quickly becomes clear is the need to understand a variety of terminology associated with operating military robotic and autonomous systems.

To decipher the above discussion, a lawyer must learn that the machine the public knows as a *drone* has different names within the military. UAV, or unmanned aerial vehicle, refers to the air vehicle itself, while UAS, or unmanned aircraft system, encompasses the entire system. UAS is a more apt description of unmanned systems as the air vehicle is just one part of the

28. Special thanks to Lieutenant Colonel Kevin Murray, USMC, a UAV commander and RAS champion, for teaching the author how to decipher UAV operator language.

entire system of systems (SoS) that includes the ground control station (GCS) and ground data terminal (GDT) that connect the UAV and GCS using a variety of datalinks or interfaces. The type of data communication between the GCS and the air vehicle depends on the UAS category or group.

The U.S. Department of Defense categorizes UAVs into five groups depending on their maximum take-off weight, ranging from less than twenty pounds to more than 1,320 pounds, and their normal operating altitude, either above ground level (AGL) or at mean sea level (MSL). Generally, the lighter the UAV is, then the lower the normal operating altitude at which it operates. See Figure 3, below, for an overview.

UAV Category	Max Takeoff Weight (lbs.)	Normal Operating Altitude (ft.)
Group 1	0-20	<1,200 AGL
Group 2	21-55	<3,500 AGL
Group 3	< 1320	< 18,000 MSL
Group 4	> 1320	
Group 5		> 18,000 MSL

Figure 3: Joint UAV Group Classifications

The smaller UAVs (Group 1 and Group 2) utilize line of sight datalinks, while the larger systems (Group 4 and Group 5) operate beyond line of sight and rely on satellite communication systems. Group 3 UAVs, such as AAI Corporation's RQ-7 Shadow platform, are larger tactical UAVs that fly slower than 250 knots. Further, UAVs may be medium-altitude long-endurance assets or high-altitude long-endurance assets, which, as these descriptions suggest, denote typical altitude and duration of the UAV mission.

The labor needed to operate a UAS is often extensive. As the UAS community jokes, "there is nothing unmanned about unmanned aviation." While the number and type of positions vary depending on the size and mission set of the UAS, common positions include the air vehicle operator (AVO), the mission payload operator (MPO), and the unmanned aircraft commander (UAC). Other specialists critical to the operation of a UAS are the intelligent specialist (IS), generator mechanics, maintenance (MX) personnel, and often the field service representative (FSR) from the contractor or manufacturer. The U.S. Air Force, perhaps recognizing the intensive labor commitment for UAS operations, calls its systems RPAs, or remotely piloted aircraft.

While this example highlights terminology common to the aviation or UAS field, learning the terminology used at sea or on the ground is equally important. The U.S. Navy has and continues to develop a variety of robotic and autonomous systems to include both UUVs (unmanned undersea vehicles) and USVs (unmanned surface vehicles).[29] The Navy has already fielded for research the *Sea Hunter*, a medium-displacement unmanned-surface vehicle, with a 132-foot hull.[30] The *Sea Hunter* is a multi-mission ship capable of operating completely autonomously on the high seas.[31] Similarly, the U.S. Army is developing an array of UGVs (unmanned ground vehicles) or unmanned ground systems (UGS) that range from small tactical systems to logistics systems capable of convoy operations.[32] Future conflict will likely include autonomous systems across the full spectrum of domains working together as a heterogeneous unit or a UXV.[33] As such, military lawyers should steel themselves for learning a variety of terminology that may or may not translate across services or domains.

Questions lawyers should to ask engineers about a particular system:

- Do you have a cheat sheet for this system's acronyms?
- Where can I read about this system's capabilities? How does it compared to ___ (a system with which I am familiar)?

III. CONTROL THEORY: HOW IT WORKS AND WHY LAWYERS SHOULD CARE

Once a lawyer grasps the basic terminology of robotic and autonomous systems, he or she can begin to understand how these systems works. Here, control and control theory are essential to understanding how these systems operate. Control, in the most fundamental sense, is simply the decision making that occurs within a system, and control theory is the mathematical and

---

29. *See, e.g.*, CHIEF OF NAVAL OPERATIONS STRATEGIC STUDIES GROUP XXVIII, WAY AHEAD PLAN: THE UNMANNED OPPORTUNITY 2 (2009), https://info.publicintelligence.net/USNavy-UnmannedOpportunity.pdf.

30. Megan Eckstein, Sea Hunter *Unmanned Ship Continues Autonomy Testing as NAVSEA Moves Forward with Draft RFP,* USNI NEWS (Apr. 29, 2019), https://news.usni.org/2019/04/29/sea-hunter-unmanned-ship-continues-autonomy-testing-as-navsea-moves-forward-with-draft-rfp.

31. *Id.*

32. *See, e.g.*, ROBOTIC SYSTEMS JOINT PROJECT OFFICE, UNMANNED GROUND SYSTEMS ROADMAP 8–10, 21–22 (2011), http://www.dtic.mil/ndia/2011/MCSC/Thompson_UGSRoadmap.pdf.

33. Thomas Pastore, George Galdorisi & Anthony Jones, *Command and Control (C2) to Enable Multi-Domain Teaming of Unmanned Vehicles (UxVs)* 1–7 (Oceans 2017 – Anchorage), https://ieeexplore.ieee.org/document/8232119.

engineering study of how automated control systems work.[34] More specifically, a control process in a system evaluates a number of variables and then selects an action to obtain the goal or desired outcome for a system.[35]

A. Control Theory: Like Colonel Boyd's OODA Loop . . . Just with More Math

The study of control theory requires a great deal of math—but before military lawyers stop reading—the essence of control theory is something taught in every professional military education curriculum: the OODA loop. Colonel John Boyd's now famous Observe-Orient-Decide-Act (OODA) process provides a great analogy to the feedback processes in control theory.[36] A system must first observe and collect data before synthesizing that data into something that it can use and make sense of. Here, one can think of this process as creating a mental map or an enabling representation. Then, in the decision phase, the system compares that representation of what is *actually* occurring to what the system knows *should* be occurring. Finally, the system acts to correct itself toward the desired goal or end state.

So how does this process look in control theory terms? First, a robot receives an input command based on real-time human input, a prior mission design, or an input from another subsystem of the robot. The controller then compares any differences or errors between this input and what the robot is doing and decides how to proceed. The processing plant (process) then executes that action (creating an output), and the robot performs a task, perhaps as simple as turning on a fan to cool its internal temperature, or as complex as deciding to alter a plan such as the original route or an intended target. During this process, the robot's sensors observe and collect data, which is then fed back into the system as another input so the controller can again decide if any adjustments need to occur (feedback).

34. MATARIĆ, *supra* note 4, at 7; FARID GOLNARAGHI & BENJAMIN C. KUO, AUTOMATIC CONTROL SYSTEMS 1–2 (9th ed. 2010).

35. CONTROL AND LEARNING IN ROBOTIC SYSTEMS vii (John X. Liu ed., 2005).

36. For an overview of John Boyd's theories and presentations, see *Introduction to the Strategic Theories of John Boyd*, THE BLASTER, http://chuckspinney.blogspot.com/p/compendium-colonel-john-boyds.html (last visited Aug. 27, 2020).

This type of control system with a feedback loop is a closed-loop control system and Figure 4 provides a visual representation of the system.[37]

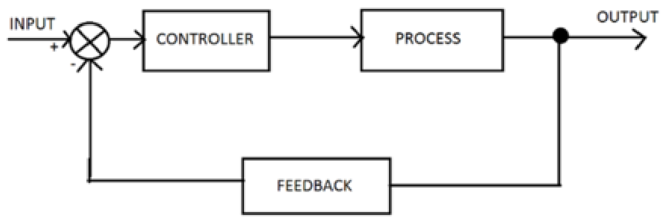

Figure 4: Basic Closed Feedback Control Loop

A complex robot will have thousands of such loops occurring simultaneously. A variety of feedback control methods are available to computer programmers (e.g., proportional, derivative, and integral), but these details exceed the scope of this article.[38]

It is important to note that when the robot's controller is making decisions, the controller is basing that decision on the representation or mapping of the external environment via its sensors, not on what is literally in the real world.[39] The sensor space or perceptual space is the full range of data that a robot can gather based on its sensors.[40] Let us take an extremely basic (and somewhat unrealistic) example of a robot with only one sensor, a thermal imaging camera. In this example, the robot's perceptual space includes the full range of temperatures the heat sensor can read, but it would not include any visual imagery, distance, or speed data. In reality, military robots have a range of sensors that often result in a larger perceptual space than that capable of the human senses (such as infrared and radiofrequency detection).

A robot then makes sense of the data it collects by developing a model, which is stored in its memory and refined via additional inputs or sensor observations. In the case of our simple temperature-sensing robot, its representation of the world would be a thermal map.

Of course, a thermal map is only one model and there are numerous ways to form or develop a model. For example, for navigation a robot may create a model that represents the actual path that it took by understanding and then remembering certain landmarks or by annotating the type of action

37. Ganesh Selvaraj, *Temperature Controlled System*, ENGINEERS GARAGE (Aug. 14, 2013), https://www.engineersgarage.com/contributions/temperature-controlled-system/.

38. MATARIĆ, *supra* note 4, at 126–31.

39. *Id.* at 23.

40. *Id.*

(e.g., turned left) that it took at certain landmarks.[41] When evaluating the model created by the robot, one also needs to know the level of uncertainty that the system has about either itself or its physical environment. Uncertainty in the model stems from the sum of any errors or "noise" in the sensors themselves and any errors or imprecision in the actions of the effectors (e.g., wheels of a ground vehicle rotated 49 degrees right instead of 45 degrees).[42] A robot's prior knowledge of the environment, such as a topographical map of the area, helps to decrease the overall uncertainty of the system.

B. *Control Architecture: How a Robot Makes Decisions*

After determining how a robot represents or models its environment, along with the system's overall level of uncertainty, the next task is to understand how the controller decides what action to take or what choice to make. While a control feedback loop that only governs one outcome (like a thermostat that turns the heat on and off at a certain temperature) is relatively straightforward, controlling multiple controllers at the same time to balance multiple—and sometimes conflicting—outcomes can become quite complex.

The rules and constraints governing these decisions are the system's control architecture.[43] Some types of control architecture allow the robot to plan its next actions by considering the outcomes of possible actions and then deciding the best alternative to achieve the desired goals or end state.[44] As the environment in which the robot operates grows in complexity and the number of possible actions increases, the planning process gets more difficult, and it takes the robot longer to determine an optimized solution.[45]

There are four basic types of control architecture methodologies: deliberative, reactive, hybrid, and behavior-based control. A deliberative approach senses the environment, plans possible actions, and then acts according to the optimized solution.[46] This approach is very methodical, taking considerable time and requiring large amount of memory storage. As such, aside from medical surgery, very few robotic systems use this approach.[47] In contrast, a reactive approach does not rely on internal representations or

41. *Id.* at 145–47.
42. *Id.* at 70–71.
43. *Id.* at 136.
44. *See id.* at 152.
45. *Id.* at 154.
46. *Id.* at 156–60.
47. *Id.* at 157.

planning but instead offers an almost reflexive approach to a given input.[48] A hybrid control architecture provides a layered mix of deliberative and reactive control depending on the particular input or subsystem used.[49] Finally, the behavior-based control architecture leads the system to take actions designed to achieve a particular behavioral outcome or goal.[50] For example, a behavior in an unmanned vehicle may be to follow a lead vehicle or not run into an object. By defining the external behavior that the system should employ, the robot then determines what actions need to happen internally to make that behavior happen.

While behavior-based control algorithms present a number of advantages for autonomous systems, the complexity increases when determining how the overall system coordinates between different behaviors. For example, an autonomous weapon system may be programmed not to fire a weapon at a child, but it also may have a behavior that allows it to fire at an individual firing at a friendly soldier. What happens if an individual is holding a child hostage while firing at a friendly soldier? Which behavior wins out?

Programmers can design a robot to arbitrate or to make a choice between two competing controls based on a priority scheme. The priority matrix could be fixed, or it could be a dynamic hierarchy that changes with various environments.[51] Alternatively, the robot could follow a cooperative control method that selects the appropriate action based on each behavior getting a "vote" in the process.[52]

The employment of lethal force requires a "reasonable" belief that the target is a legal one.[53] To give sound legal advice concerning the legality of the employment of lethal autonomous weapon systems, military lawyers must understand both the representation models used by those systems and the type of control approaches utilized. Reasonableness ties directly to how an autonomous robot creates and then evaluates representations of its perceptual space. In other words, will the autonomous robot create a model of a given target and differentiate the target (combatant versus civilian, military versus civilian object) to the same reasonableness standard as a human being.

---

48. *Id.* at 161–76.

49. *Id.* at 177–86.

50. *Id.* at 187–205.

51. *Id.* at 207–08. This arbitration method of resolving conflicts between various control systems is also known as a competitive control method.

52. *Id.* (noting that programmers also refer to this method as behavior fusion).

53. Michael N. Schmitt & Jeffrey S. Thurnher, *'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict*, 4 HARVARD NATIONAL SECURITY JOURNAL 231, 257 (2013).

Likewise, target selection necessarily requires an autonomous system to make decisions based on desired behaviors, which may conflict with one another. To ensure the legality of such decisions, a military lawyer must understand how a robot will arrive at an optimized action and whether the robot will resolve conflicts via a competitive or cooperative control method.

Questions lawyers should to ask engineers about control systems:

- What is the system's perceptual space?
- What type of control algorithms does this system use and why?
- What type of behaviors does the system want to achieve?
- How does the system resolve conflicts between behaviors?

IV. AUTONOMOUS THINKING: HOW ROBOTS LEARN AND REASON

If an autonomous system can make choices in deciding how to accomplish a particular behavior or a specific outcome, the next question is whether a robot can make those choices or achieve these outcomes better or faster the next time.[54] In other words, can the system learn? If so, how does it reason its way to a decision? To determine whether an autonomous system can make reasonable decisions under the law of armed conflict, the military lawyer must also understand both how a system learns and how the robot reasons its way to a decision.

A. Learning Machine Style

Learning occurs when the robot can accomplish the goal or task faster or with improved accuracy in subsequent attempts.[55] A robot also learns whenever it understands and then uses previously unknown information (e.g., to create a map and then apply the map in making a decision) or when it understands and then acts in a changing environment.[56] Put simply, the robot is accomplishing a task based on information it did not have before and using

54. Rodney A. Brooks & Maja J. Matarić, *Real Robots, Real Learning Problems*, *in* ROBOT LEARNING 193 (Jonathan H. Connell & Sridhar Mahadevan eds., 1993) (asking whether a robot can be "programmed to do more than it was programmed to do").

55. BEKEY, *supra* note 5, at 125–26.

56. *Id. at* 126.

that information in a way or manner for which it was not specifically programmed. When this learning occurs within a computer, it is called machine learning. In contrast, robotic learning occurs as a robot interacts with the world or the external environment.[57] Both of these discrete fields of study fall under the more global research area of artificial intelligence.[58]

Software engineers have developed a range of approaches to enable systems to learn, although robotic learning is still in the early stages of development. The approaches and design techniques vary according to what the engineers need the system to learn.[59] For example, does the robot need to learn something about its environment—either internal or external—or does the robot need to learn a new skill or behavior? Regardless of the type of learning necessary to accomplish, an engineer must account for three basic attributes of the learning process: performance feedback, memory, and training.[60] Thus, a robot must be capable of evaluating its performance to learn from it (performance feedback). The system must then have a means of storing such information for its future use (memory).[61] Finally, training is the mechanism that links the lessons from the evaluation process to create knowledge in the robot's memory.[62]

One common approach is reinforcement learning. Here, the robot attempts new actions and receives a reward (positive feedback) for completing a positive action or a punishment (negative feedback) for taking an action inconsistent with the desired outcome.[63] One of the challenges with reinforcement learning is how the robot assigns "credit" for actions that lead to reward versus those actions that lead to punishment.[64] Not every decision leads to an immediately recognizable outcome. As a result, a robot may not credit an action separated by either time or space even though the action

57. *Id.* at 128.

58. JANET FINLAY & ALAN DIX, AN INTRODUCTION TO ARTIFICIAL INTELLIGENCE 77 (1996). One roboticist defines artificial intelligence as "the science of endowing programs with the ability to change themselves for the better as a result of their own experiences." RONALD C. ARKIN, BEHAVIOR-BASED ROBOTICS 305 (1998).

59. *See* ARKIN, *supra* note 58, ch. 8; *see generally* MAKING ROBOTS SMARTER: COMBINING SENSING AND ACTION THROUGH ROBOT LEARNING (Katharina Morik, Michael Kaiser & Volker Klingspor eds., 1999).

60. Jay Farrell & Walter Baker, *Learning Control Systems: Motivation and Implementation*, *in* INTELLIGENT CONTROL SYSTEMS: THEORY AND APPLICATIONS 443, 454 (Madan M. Gupta & Naresh K. Sinha eds., 1996).

61. *Id.*

62. *Id.*

63. MATARIĆ, *supra* note 4, at 256–61; Brooks & Matarić, *supra* note 54, at 206–07.

64. MATARIĆ, *supra* note 4, at 259–60.

contributes to the final positive result. Another approach is supervised learning, where an external "teacher" (e.g., learning algorithms) tells the robot whether its actions or decisions were good ones or not.[65] These and other robotic learning approaches have advantages and disadvantages depending foremost on what the robot needs to learn and the environment in which the robot is operating.

Questions lawyers should to ask engineers about robotic learning:

- What type of learning is this system designed to do?
- How does the system learn? What check does the human have on whether the system has learned proper lessons?
- How does the system adapt to a changing battlefield? Does the human operator make these adaptations through communications, or is the system capable of recognizing the battlefield has changed?

B. Robot Reasoning

A concept related to the idea of learning is that of reasoning and a critical understanding for military lawyers to have about autonomous systems is the distinction between inductive and deductive reasoning.[66] Inductive reasoning extracts general principles from specific instances.[67] Inductive reasoning allows a solution to be generated by a hunch or a guess (an educated one or just dumb luck), which is then tested in the real world to see if it is correct.[68] Humans are extremely good at inductive reasoning and regularly use this approach to solve problems. Each time a doctor makes a diagnosis, he or she uses inductive reasoning. A lawyer relies on inductive reasoning each time he or she makes a conclusion based on various pieces of evidence or sets out a theory of the case based on a series of assumptions.

65. *Id.* at 260–61.

66. For a detailed discussion of reasoning models and their ethical implications in autonomous systems, see Scott Heritsch & Linell A. Letendre, *Engineering Model for Ethical Decision-Making and Regulation in Autonomous Systems*, *in* HANDBOOK OF UNMANNED AERIAL VEHICLES (Kimon P. Valavanis & George J. Vachtsevanos eds., 2018).

67. BEKEY, *supra* note 5, at 184.

68. Steven D. Harris & Jennifer Narkevicius, *Appropriate Automation: Human System Dynamics for Control Systems*, CONTROL ENGINEERING (Feb. 28, 2014), https://www.controleng.com/articles/appropriate-automation-human-system-dynamics-for-control-systems/.

In a simple battlefield example, if a soldier in a counterinsurgency environment sees a specific truck drive down the same road at the same time for five consecutive days, stop, wait ten minutes, and then leave, the soldier may predict that that same truck will come down the same road tomorrow. If the truck does return the next day, but the driver abandons the vehicle by running away from it, the soldier may assume that the truck contains explosives. Here, the soldier induced a general conclusion from contextual clues and a guess about the driver's intent. Of course, the conclusion may be incorrect. The driver may be making a routine delivery each day and in this instance, needed to abandon the vehicle for a variety of reasons. But if correct, the soldier's inductive reasoning skills could allow him or her the opportunity to alert others to this threat.

In contrast, deductive reasoning starts with a general theory or premise and then uses a top-down approach to determine if a specific fact or situation is true. Deductive reasoning relies on existing facts that the system knows (or is told) to be true and then determines whether the new, more specific situation is true.[69] Computers (and as a consequence robots) are extremely good—and often much better than humans—at deductive reasoning.[70] This is because they can retain an incredible amount of facts and data and then use the closed-loop feedback system repeatedly until it solves the question.[71] Thus, computers can solve extraordinarily complicated math problems much faster (and without error) than humans can. Likewise, a UAS can land on a carrier deck in the exact same spot each time.[72]

So why do lawyers need to care about inductive versus deductive reasoning? Because, as Harris and Narkevicius note, engineers cannot yet automate induction.[73] Thus, a fully autonomous weapons system is not technically feasible, and human-interface is not just a good idea, it is necessary to the control process.[74] As explored in the companion article for engineers,[75] the law of armed conflict requires sound judgment and reasonable interpretations as

69. FINLAY & DIX, *supra* note 58, at 82.

70. Harris & Narkevicius, *supra* note 68.

71. *See supra* Part III.A.

72. During initial testing, the Navy's X-47 UCLASS (Unmanned Carrier Launched Airborne Surveillance and Strike System) landed in exactly the same spot so many times it left marks on the carrier deck. Interview with Navy UCLASS Testing Personnel at Naval Air Station, in Patuxent River, MD (Feb. 5, 2015).

73. Harris & Narkevicius, *supra* note 68.

74. *Id.*

75. *See supra* note 1.

to the intentions of adversaries on the battlefield.[76] Robots—even amazing supercomputers like IBM's Watson[77]—currently are incapable of such reasoning. As Harris and Narkevicius explain, "[i]ntentionality, by its very nature, cannot be observed; it can only be inferred; that is, induced."[78]

The second reason military lawyers need to understand reasoning is to ask tough questions about how information and decision making is shared between human operators and autonomous systems. The U.S. military seeks to achieve this balance by applying a concept called "manned-unmanned teaming."[79] This concept follows a "system-of-systems" approach and rather than attempting to develop a robotic system that can "do it all," it divides tasks best suited to humans to human operators and those best suited for machines to robots.[80]

Still, to make good inductive judgments humans need reliable and actionable information. Thus, as machines get increasingly better at making decisions and human operators move from *in-the-loop* control to *on-the-loop* oversight (and at some point likely *on-multiple-loops* oversight),[81] practitioners must understand what information is shared with the human operator to ensure that he or she can make sound, informed inductive decisions. Otherwise, as robotic systems become faster, humans risk becoming the weakest link in the decision-making process chain and meaningful human control becomes a fanciful term.[82]

---

76. For additional discussion of the incorporation of artificial intelligence and machine learning in weapon systems, see Alan L. Schuller, *At the Crossroads of Control: The Intersection of Artificial Intelligence in Autonomous Weapon Systems with International Law*, 8 HARVARD NATIONAL SECURITY JOURNAL 380 (2017).

77. *See IBM Watson Supercomputer*, TECHOPEDIA, https://www.techopedia.com/definition/15347/ibm-watson-supercomputer (last updated Aug. 18, 2017) ("IBM's Watson supercomputer is a question-answering supercomputer that uses artificial intelligence to perform cognitive computing and data analysis.").

78. Harris & Narkevicius, *supra* note 68.

79. THE DWIGHT D. EISENHOWER SCHOOL FOR NATIONAL SECURITY AND RESOURCE STRATEGY, NATIONAL DEFENSE UNIVERSITY, FINAL REPORT: ROBOTICS AND AUTONOMOUS SYSTEMS 10, 12, 18 (2015), http://es.ndu.edu/Portals/75/Documents/industry-study/reports/2015/es-is-report-robotics-autonomous-systems-2015.pdf.

80. *Id.* at 10.

81. *See* Schmitt & Thurnher, *supra* note 53, at 235–37.

82. NICHOLAS CARR, THE GLASS CAGE: AUTOMATION AND US 191 (2014).

Questions lawyers should to ask engineers about reasoning:

- What types of information and deductive decisions does the system share with the human operator?
- For what type of inductive decisions does the system rely on the human operator?

V. CONCLUSION

> *The first shot freely taken by a robot will be a shot heard round the world. It will change war, and maybe society, forever.*[83]

While robotics engineers are the first to admit that the ability of a fully autonomous weapon system to identify, select, and then engage an enemy target without human intervention is still some time in the distant future, the need to understand how autonomy and international law coincide and interact exists today. For military lawyers to provide sound advice on this topic, they cannot simply have a general overview of a system's capabilities. Rather, they must immerse themselves in the technical details of why robots work the way they do and what technical limitations autonomy possesses. Likewise, engineers must embrace at least a functional understanding of international legal principles to develop capable systems that comply with the law.

Noted scientist and philosopher Sir Francis Bacon once said, "The inquiry, knowledge, and belief of truth is the sovereign good of human nature."[84] For all the debate of "killer robots,"[85] and the legal and ethical implications of employing lethal autonomous weapon systems,[86] I remain confident that humankind will find an acceptable path—both legally and ethically—for their use. When engineers, lawyers, and commanders work together and communicate in the same language, they will spark the inquiry and examination necessary to find solutions to the considerable challenge created by rapidly advancing technological capabilities. And in the end, they just might get through those pearly gates together.

---

83. *Id.* at 193.

84. Sir Francis Bacon, *Of Truth*, *in* ESSAYS OR COUNSELS, http://fly.hiwaay.net/~paul/bacon/essays/truth.html (last visited Aug. 27, 2020).

85. ARMIN KRISHNAN, KILLER ROBOTS: LEGALITY AND ETHICALITY OF AUTONOMOUS WEAPONS (2009). *But see* Ronald C. Arkin, *The Moral Case for Autonomy in Unmanned Systems*, *in* HANDBOOK OF UNMANNED AERIAL VEHICLES, *supra* note 66, at 2933–41.

86. Charles J. Dunlap, Jr., *Accountability and Autonomous Weapons: Much Ado about Nothing?*, 30 TEMPLE INTERNATIONAL AND COMPARATIVE LAW JOURNAL 63 (2016).