OUT OF THE LOOP

By Gary Brown*

There has been considerable discussion of and consternation over the possibility of developing autonomous weapon systems (AWS). Some of the rhetoric has been splendidly excessive—“killer robots,” “the end of the human race,” etc. Although many of the objections come from individuals with legal backgrounds, they go beyond legal concerns to discuss the morality or ethics of autonomous machines taking human life. In this article, I suggest that opposition to AWS is not especially useful, both because they already exist and will continue to be developed and used whether we like it or not (i.e., “resistance is futile”) and because machines might be better suited to the task than humans anyway (i.e., they might be exactly “the droids you’re looking for”).

It is human nature to be skeptical of new ways of doing things, but technological advancement is normal. Some innovation fails, while other ideas become so ingrained in everyday life that it is difficult to remember when we did not have them. For example, in the nineteenth century it would have been unthinkable for an ordinary rider to “drive” an elevator. A trained operator ensured that passengers were safely transported between floors. The driverless elevator was developed by 1900, but public trepidation over the seemingly zany idea of self-service elevators prevented the deployment of the technology until

* Gary Brown is Professor of Cyber Security at Marine Corps University. Previously, he served as Head of Communications for the Washington Delegation of the International Committee of the Red Cross. He was the senior legal advisor for U.S. Cyber Command for three years, and is a retired U.S. Air Force colonel.

1. For purposes of the article, I will use the Department of Defense’s definition: “A weapon system that, once activated, can select and engage targets without further intervention by a human operator.” U.S. Dept. of Def., Directive No. 3000.09, Autonomy in Weapon Systems 1, 13 (Nov. 21, 2012).


1945—and even then it took an elevator operator strike in New York City to force the issue. Now, it is hard to imagine that driverless elevators were the self-driving cars of post-WWII America. Like the elevator, it will not be too many years before it seems silly that we ever thought the artificial intelligence (AI) operating cars would be less safe than human drivers. These drivers, who might be skilled and attentive, might also at any given time be unskilled and/or intoxicated, texting, applying lipstick or arguing with a passenger, for example.

As the technology to develop AWS has progressed, opposition has also progressed. Many objections to AWS are based on the position that it would be amoral to remove humans from the decision to use lethal force. The positions of three organizations on the appropriateness of AWS are discussed below.

A 2015 Human Rights Watch (HRW) report, entitled Mind the Gap: The Lack of Accountability for Killer Robots, raised several moral and legal concerns with AWS. HRW asserted that the decision to kill a human being should not be left to a machine because it violates victims’ dignity. The report also argued the technology may not be able to replicate human judgment in lethal situations and noted the possibility that AWS could spark a global arms race.

HRW’s primary assertion is that AWS may violate the law of armed conflict (LOAC), within international humanitarian law, because programming flaws could result in LOAC violations that might have been avoided if a human were in the decision loop. The HRW report notes humans are deterred from criminal behavior by the possibility of legal consequences, making a human-guided lethal system more compliant. Taking the legal argument a step further, HRW suggested there is insufficient legal accountability for criminal or negligent acts during the production or operation of AWS.

On a more futuristic trajectory, over a thousand robotics researchers signed a letter opposing the development of AWS; it was released at the 2015 International Joint Conferences on Artificial Intelligences. The signatories included names as
eminent as Elon Musk, Stephen Hawking, and Steve Wozniak.\(^\text{18}\) These researchers also noted the possibility of an arms race, and discussed various unsavory possible uses for AWS, such as ethnic cleansing,\(^\text{19}\) which seems to be a bit of a red herring. More interestingly, they expressed concern that the AI developed to allow AWS to make independent targeting decisions might slip out of humanity’s control as the systems become too complex for humans to master.\(^\text{20}\)

The International Committee of the Red Cross (ICRC) argues AWS are not yet capable of complex decision-making and reasoning, and that increased autonomy will lead to increasing unpredictability in the way these lethal weapons will operate.\(^\text{21}\) This objection suggests that AWS might not be in compliance with LOAC, but the ICRC does not stop there.\(^\text{22}\) It also notes that using lethal AWS would demonstrate a lack of respect for human dignity and would generally be neither ethically nor morally acceptable.\(^\text{23}\) The ICRC is not currently suggesting a ban or moratorium on AWS, but is advocating for more consideration of the issues it raised before developing or using these weapons.\(^\text{24}\)

The three efforts summarized here object to AWS based on under-sophisticated technology, overly sophisticated technology, the potential of a new arms race, for moral or philosophical concerns, and on legal grounds. I will leave it to others to discuss whether the technology is too cold, too hot, or just right. I will also not address a potential new arms race; as argued below, the technology is already extant, and will probably proliferate in a predictable manner. Besides, the issue of a possible arms race is not unique to AWS technology.\(^\text{25}\)

The moral issue of a machine taking a human life, if answerable at all, will have to be addressed by philosophers, not lawyers. Even though the Department of Defense’s 2015 Law of War Manual (Manual) discusses the principle of honor,

---

\(^\text{18}\) Id.
\(^\text{19}\) Id.
\(^\text{20}\) See id. (noting that autonomous weapons select and engage targets without human control).
\(^\text{22}\) Id. at 2.
\(^\text{23}\) Id.
\(^\text{25}\) See Autonomous Weapons: An Open Letter from AI & Robotics Researchers, supra note 17 (comparing the potential arms race of autonomous weapons to previous Kalashnikovs arms race).
which seems to have a moral rather than a legal basis, it remains grounded in law.\textsuperscript{26} The Manual notes that honor “requires a certain amount of fairness in offense and defense [and] forbids resort to means, expedients, or conduct that would constitute a breach of trust with the enemy.”\textsuperscript{27} The referenced “trust” seems to be the level necessary to make laws of war relevant.\textsuperscript{28} It is not clear that using a machine to apply lethal force is any more a breach of trust than using long bows to penetrate armor or artillery to strike targets over the horizon. Whether it is a long-range arrow or a 155 mm round, at the time of attack there is no opportunity for the target to communicate with the human agent,\textsuperscript{29} whether to plead for compassion or indicate the attack is somehow inappropriate. Considering the morality of ranged weapons as a class might be interesting, but at least to this lawyer, there does not seem to be a unique moral issue presented by moving humans one additional step back from the trigger pull.

The legal objections to AWS resolve into two general categories: insufficient legal accountability for LOAC violations and a level of unpredictability that somehow violates LOAC. These legal concerns may best be addressed by examining the similarities and differences in AWS and humans exercising lethal force. One similarity is that machines must be programmed—but humans must be trained, and there is room for error in both processes. Another similarity is that outside actors may cause both to behave in undesirable ways. All AWS can, in theory, be hacked. In fact, humans can be tricked, turned, blackmailed or otherwise convinced to act against rules of engagement (ROE).\textsuperscript{30}

One difference between machines and humans is that machines have no morality.\textsuperscript{31} This is generally cited as a disadvantage of AWS because humans are presumed to have some moral compass.\textsuperscript{32} Unfortunately, what constitutes morality appears to vary widely.\textsuperscript{33} Some humans in a given population could be expected to have a distorted morality that allows them to engage in murder, assault and other mayhem, which can hardly be seen as an advantage. Opponents of AWS hope that humans will show mercy in certain combat situations, even if ROE are met to use lethal force, and thereby save lives.\textsuperscript{34} Certainly, this can happen.\textsuperscript{35} Sadly, it is also


\textsuperscript{27} Id. at 66.

\textsuperscript{28} See id. at 64 (“[A] breach of trust . . . undermines respect for the law of war.”).

\textsuperscript{29} See id. at 66 (finding that hostilities are not unlimited and discretion is required to uphold honor, fairness, and trust between combatants).

\textsuperscript{30} See id. at 26–27 (defining rules of engagement as guidelines States issue to their militaries to implement the law of armed conflict and other restrictions they choose to impose).

\textsuperscript{31} See generally Asaro, supra note 10.

\textsuperscript{32} See id. at 708 (discussing the objection to AWS on the basis of an absence of a moral agent).

\textsuperscript{33} Id.

\textsuperscript{34} See id. at 698–99 (explaining the multiple layers of interpretation and judgment in following guidelines to determine whether subject is a legitimate target).
possible that humans will be motivated by anger, fear, anxiety or other negative emotions to violate ROE in an aggressive fashion. Humans might even then lie about their actions to cover up the misconduct. AWS are immune from these issues.

Another important difference is that on one hand, machines cannot die, so they do not fear death and can be programmed to accept high levels of risk. Humans, on the other hand, can be killed, and most sane ones would prefer to live. AWS lack of instinct for self-preservation could be used to structure ROE that would better protect civilians. For example, AWS can be directed through programmed road checkpoint ROE to shoot to wound, or shoot out tires or engine blocks, even though these are less effective than shooting to kill, because the result of failing to stop a car packed with explosives would only be destroyed AWS, not dead human guards.

Concerned lawyers often cite the lack of meaningful accountability for machines as a reason to prohibit lethal AWS. The argument is that if there is a LOAC violation involving AWS there may be no one who can be held accountable in a meaningful way. Their presumption is that this is different than situations involving humans in the loop of lethal weapon systems. In cases involving humans in the loop, there might be an investigation of the actor and the chain of command. If ROE are found to have been flawed, or ordered to be wrongly applied, there can be liability for superior officers. That could also be the case for AWS. If ROE were overly permissive, the authors of the ROE can be held accountable.

---

35. See generally Docherty, supra note 2.
36. See id. (considering the potential for humans to be tempted in the heat of battle to misuse AWS and increase the risk of violating the laws of war).
37. See id. (recognizing the possibility that humans could attempt to avoid responsibility of deploying an automated weapon that committed a criminal act by blaming others).
38. See id. (stating that fully autonomous weapons lack certain human characteristics like emotion and intentionality). SciFi fans may appreciate that AWS objectors view it as Captain Picard vs. The Borg; I am suggesting it might rather be Khan (and all his Wrath) vs. Spock.
39. See Asaro, supra note 10, at 692, 705 (assessing the risk redistribution in the use of AWS).
40. See id. at 696 (discussing the ability to program AWS to have them sacrifice themselves in situations where humans would not do the same).
41. Id.
42. See Docherty, supra note 2 (analyzing the lack of accountability of wrongful action crimes by autonomous weapons).
43. Id.
44. Id.
45. See id. (considering direct and command responsibility in the commission of wrongful acts).
46. See id. (discussing the possible liability for the operator or the commanding officer in the event of criminal acts perpetrated by the autonomous weapon).
responsible. In AWS, ROE would only be wrongly applied if the system were programmed incorrectly. Why would those responsible for programming not be liable? If the AWS were manufactured incorrectly, manufacturers could be held just as liable as those who manufacture bombs, rockets, bullets or other traditional weapons that malfunction. The only real difference in accountability is that with AWS there is no human trigger-puller to hold accountable. The tradeoff is that AWS will do what they are told, without regard to fear, anger, hunger, exhaustion, or simple maliciousness. Additionally, AWS can accept high levels of risk to themselves because they have no life to preserve, so ROE can be written to err on the side of caution up to the last possible moment.

With their dispassionate application of ROE while providing an increased acceptance of risk, AWS offer the potential to be more effective than human agents in protecting noncombatant lives during armed conflict. If the only reason to introduce an additional, potentially flawed actor into the equation is so there is someone else to blame when things go wrong, it is a thin justification.

Suggestions of the superiority of human judgment to that of machines might be as misguided as the arguments about the legal superiority of a regime without AWS. In complex situations, that is not always the case. Autonomous systems designed to be dispassionate in complicated situations may perform better than humans. For example, two high-profile plane crashes might have been avoided if the pilots had relied on the autopilot rather than acting as they thought best. Air France Flight 447 crashed and killed all 228 people aboard while, en route from Rio de Janeiro to Paris, the pilots disengaged the autopilot, preventing it from liabilities for wrongful criminal acts of AWS).

48. See Asaro, supra note 10, at 696 (analyzing the ability to program AWS to conform to specific ROE).

49. Which is to say, in all honesty, not very liable. See Docherty, supra note 2 (discussing the difficulty of using civil liability mechanisms to hold a manufacturer of autonomous weapons accountable).

50. “The implied responsibility placed upon the officer’s shoulders by the subordinate’s unhesitating willingness to follow orders is a withering burden to any officer with half a brain . . . .” NEAL STEPHENSON, CRYPTONOMICON 113 (1999). Maybe a concern is creating a system that actually will do precisely what it is told.

51. See id. (explaining the benefits of AWS).

52. Id.

53. See U.S. Dept. of Def., Directive No. 3000.09, supra note 1, at 1 (establishing DoD policy for the development and use of autonomous and semi-autonomous functions in weapons systems).

54. See id. (discussing situations where AWS are more accurate than human judgment).

55. See id. (comparing AWS to humans).

taking the counterintuitive actions that might have saved the plane.\textsuperscript{57} We cannot know for certain if the autopilot would have been able to save the plane, but clearly the pilots’ actions made the situation worse—\textsuperscript{58} a common result of a system being so complicated that humans are not able to keep up with what the computer is doing.

In the case of Asiana Airlines Flight 214, a complex algorithm that the pilot apparently did not understand chose an autopilot mode that guided the plane, but still required manual throttle control.\textsuperscript{59} The pilot failed to recognize this requirement, and, as a result, failed to provide the necessary input.\textsuperscript{60} The plane’s tail hit a seawall while landing in San Francisco, resulting in three deaths and 187 injuries.\textsuperscript{61} Without a human pilot, the autopilot would not have had the manual throttle option.\textsuperscript{62} Once again, in this case, having a human in the loop caused the accident.\textsuperscript{53}

In both of these cases, it would be easy to conclude that the fault lies with the system designer for making a system so complex a human cannot understand it. The problem with that conclusion is that it ignores that automation can exceed human capabilities, and can therefore increase reliability and safety.\textsuperscript{64} It seems that, at the moment, automated flight is not perfect enough to be relied on without human supervision, but is becoming so complicated that human intervention is

\textsuperscript{57} The airspeed indicator on the plane had iced over and provided erroneous data, which fatally confused the aircrew, and the crash investigation suggested the autopilot would have outperformed the crew in trying to recover the plane. See Bruton, \textit{supra} note 56 (detailing the events that occurred leading up to the crash of Air France 447); see David Learmount, \textit{IN FOCUS: After AF447, What Now?}, FLIGHTGLOBAL (Oct. 17, 2012), https://www.flightglobal.com/news/articles/in-focus-after-af447-what-now-377433/ (discussing the final accident report on AF447’s causes, which stated that the pilots’ reaction based on their mental representation of the situation rather than the real situation resulted in the loss of control that could have been recovered if standard procedures had been followed).

\textsuperscript{58} See Bruton, \textit{supra} note 56 (explaining how the pilot made the situation worse by pointing the airplane’s nose up instead of pointing it down which may have prevented the plane from crashing).

\textsuperscript{59} See Asiana Airlines Crash Caused by Pilot Error and Confusion, Investigators Say, \textit{supra} note 55 (detailing the circumstances leading up to the Asiana Airlines crash).

\textsuperscript{60} See id. (describing how the pilot’s failure to understand the plane’s automated system led to the resulting plane crash).

\textsuperscript{61} See id. (detailing the plane crash).

\textsuperscript{62} See id. (explaining how the plane crash could have been avoided).

\textsuperscript{63} By offering these two examples, I do not mean to suggest that heroic efforts by aircrew never prevent airplane crashes; they do. Once we have introduced complex automation, however, it is almost certain we will continue improving it until it is superior to human skill all, or nearly all, the time. As noted here, the transition phase can be especially dicey, as parameters and rule sets are reformed to fit computer capacity, leaving poor human ability to calculate in the dust.

increasingly less effective. The same situation might be said to exist regarding LOAC.

In an effort to increase the morality and decrease the destruction of war, we have expanded the law that applies to warfare dramatically. ROE and the implementation of these laws have also increased in length and complexity and are supplemented by special instructions and other documents that often run to hundreds of pages. As lethal engagements may occur in the space of seconds or minutes, it is increasingly less reasonable to expect combatants to apply complex rules that puzzle even scholars. When it comes to applying long lists of rules in a short period of time, computers are superior to humans. If the determination is that long lists of rules and caveats are necessary, and that observing them will result in lawful actions, it follows that the entity best at applying them—AWS—must be lawful. What would be the point of designing a system to ensure legality, while at the same time prohibiting the tool that would best enable compliance?

Even if we are to conclude that, despite being lawful, AWS should not be developed, it is too late because AWS already exist. Today, autonomous machine guns are capable of finding, tracking, warning, and eliminating human targets, without human oversight. In fact, a South Korean company has already developed Super aEgis II, an automated machine gun that can track and eliminate targets without any human intervention.

Nevertheless, States could pursue a ban on further development and proliferation of AWS. It is unlikely such an effort, even if implemented, would have much of an effect. It might prevent activity in the area by law-abiding States, but would do little to stop the development of AWS by noncompliant States or

65. See id. at 1-8 to 1-10 (describing the complexities of automated flight and how human intervention is becoming increasingly less effective).
67. See id. (discussing the length and complexity of the laws of war).
68. Cf. Thomas K. Adams, Future Warfare and the Decline of Human Decisionmaking, PARAMETERS, Winter 2001–02, at 5 (stating that the concept of information warfare favors continually increasing the flow of detailed information, but human decision-making capacity will invariably fail to keep up).
69. This is redolent of recent debates about the use of instant replays to help officials in National Football League games. History of Instant Replay, NFL, http://operations.nfl.com/the-game/history-of-instant-replay/ (last visited Feb. 29, 2016). As football rules became more complicated, it became obvious that officials needed technical assistance to make correct calls. Id. The (losing) argument against video reviews came down to preserving tradition, which is an even poorer argument regarding compliance with the law (there was also a speed of play argument against replays that is not relevant here). Id.
70. See ICRC REPORT ON AUTONOMOUS WEAPON SYSTEMS, supra note 21, at 1 (urging States to consider reaching an agreement on what makes human control of a weapon adequate).
72. See id. (describing the features of the Super aEgis II).
armed groups seeking an asymmetrical advantage in warfare. Unlike weapons of mass destruction, AWS require no sensitive materials or even much unique technology. Researchers are developing technologies, such as software to identify people by face or gait, and AI, for other purposes that would be useful in AWS. Because the technology and materials are or will become generally available, policing a ban would be extremely difficult. The United Kingdom (U.K.), for one, seems to understand the futility of an AWS development ban. At a recent United Nations conference in Geneva on the future of weaponry, the U.K. opposed a ban on the development of AWS.

In one sense, because AWS are already available, it is perhaps too late for a ban. However, in another sense, it might be too early for a ban. After being used, some weapons have been found to be so terrible that an international consensus builds to ban or control them. The reason the Non-Proliferation Treaty and the Chemical Weapons Convention have been largely successful is that people witnessed the tragic results of using these weapons. However, AWS have not been used in conflict in any significant way. There are not warehouses full of

73. See Tia Ghose, Ban Killer Robots Before They Take Over, Stephen Hawking & Elon Musk Say, LIVESCIENCE (July 27, 2015), http://www.livescience.com/51664-stephen-hawking-elon-musk-ai-weapons.html (stating that a global arms race to make artificial-intelligence-based autonomous weapons is almost sure to occur unless nations can ban the development of such weapons).


75. See Owen Bowcott, UK Opposes International Ban on Developing ‘Killer Robots’, THE GUARDIAN (Apr. 13, 2015), http://www.theguardian.com/politics/2015/apr/13/uk-opposes-international-ban-on-developing-killer-robots (quoting the Foreign Office’s statement that there is no need for outlawing the autonomous weapon because international humanitarian law already provides sufficient regulation for this area).

76. See id. (stating that technical judges have all the rights and duties of a professional judge).

77. See Genesis and Historical Development, ORG. FOR THE PROHIBITION OF CHEMICAL WEAPONS, https://www.opcw.org/chemical-weapons-convention/genesis-and-historical-development (last visited Feb. 23, 2016) (stating that chemical weapons have always been viewed as particularly abhorrent and that there was a long history of international agreement of limiting the use of the chemical weapons).

78. See id. (explaining the history and background of the Chemical Weapons Convention); see also Daryl Kimball, The Nuclear Nonproliferation Treaty (NPT) at a Glance, ARMS CONTROL ASS’N, (Aug. 2012), https://www.armscontrol.org/factsheets/nptfact (stating that with its near-universal membership, the NPT has the widest adherence of any arms control agreement).

new, effective weapons that have never been fielded.\(^{80}\) There may be budgetary or strategic reasons not to deploy weapons, but if they work, eventually they will be put to use. If there is a sweet spot for negotiating weapons bans, it is only after they have been employed and a consensus on the need to control them follows.\(^{81}\)

The AWS opponents discussed here focus on two aspects of employing AWS, suggesting that only people should kill people and that when something goes wrong, somebody ought to be able to be held responsible. As discussed above, these objections provide insufficient grounds to ban AWS.

Ultimately, we may decide that humans should determine the lethal use of weaponry should on every occasion. However, there is currently too little evidence to conclude that warfare, as currently conducted, is morally and legally superior to how it might be conducted with the incorporation of a new generation of machines. AWS are not prohibited by LOAC, and vague philosophical concerns should not—and will not—prevent their development and fielding. The only question is whether the United States and its allies will remain on the sidelines or participate in the process of developing the weapons and rules regulating their use.

---

80. See id. at 1030 (asserting that it is unrealistic that States refrain from deploying autonomous weapons).

81. This argument leaves aside weapons that are clearly beyond the pale of reason, including those that would extinguish all human life, for example. We can be fairly certain Star Trek’s Genesis Device, designed to reduce everything on a planet to subatomic particles, would be banned ab initio.