Policy Memorandum to the Emerging Threats and Capabilities Subcommittee of the Senate Armed Services Committee: Securing our Future Against the Threat of Lethal Autonomous Weapon Systems

Brian Fydenkevez

United States Military Academy
Department of Physics and Nuclear Engineering, West Point, NY 10997
Corresponding author: Brian.J.Fydenkevez.mil@mail.mil

Executive Summary:
The first serviceable autonomous robots may soon command and occupy our military battlefields. These robots will kill humans and destroy enemy property based on algorithms and internal calculations, without the direct guidance of a human operator. The decision to employ these robots will shift the way we wage warfare, virtually removing friendly military forces from the battlefield. Doing so comes at a long-term cost to both foreign and U.S. civilians as the nation's adversaries contemporaneously develop these same technologies. These weapons face insurmountable barriers to maintaining *Jus in Bello*, the international humanitarian law that regulates conduct of belligerents during wartime in an effort to prevent undue suffering. They present significant shortcomings with regard to identifying targets and ensuring accountability in warfare. Consequently, I call on Senator Debra Fischer to (1) ban the development and use of these weapons by the United States and (2) advocate for a similar global ban.

I. A New Warfare

Lethal autonomous weapons systems (LAWS) take a human out of the process of killing. These robots make the determination to kill without any input from a person, executing a series of algorithms that consider the input from their signal intelligence and yield an output with the potential to take human life. While killer robots may seem the topic of a dystopian future, the technology already exists for modern robots to fill the role. In fact, over five years ago, a Korean company designed and deployed a computerized machine gun that can identify and target humans who enter the Demilitarized Zone.[i] With a slight modification to its code this robot could also shoot a particular target without ever asking a human commander if doing so is the appropriate choice.

Rebecca Crootof delineates four categories of weapons based on decreasing human control. Inert weapons, including swords or guns, require human operation in order to be lethal. Automated weapons react to stimuli, like the aforementioned robot or a landmine, responding to a trigger rather than actually weighing data. Semi-autonomous and autonomous weapon systems may select and engage targets, but the former can do so only while a human operator takes some affirmative action to select a specific target

while the latter internally and independently perform all functions germane to killing.[ii] For the purposes of this policy recommendation, LAWS will refer to autonomous weapon systems as defined by Crootof, meaning that the robot is capable of killing without action required on the part of a human operator.

As nations continue to develop these new technologies, the advantages of autonomous robots will drive states to pursue their development. Robots can go where humans cannot, work longer, maintain alertness, and perform consistently across varied situations without complaint. For these reasons, the U.S. Air Force predicted, "by 2030 machine capabilities will have increased to the point that humans will have become the weakest component in a wide array of systems and processes."[iii]

II. A Dangerous Course

Although sending robots capable of exercising lethal force autonomously in the place of human combatants seems like a means to minimize the suffering that warfare causes, the threat that such weapons pose in their long-term development and use outweighs the immediate military advantage they might offer to a nation that uses them. Employment of LAWS presents two major unique and irresolvable problems. The first of these is distinction–the efficient and correct identification of targets. Because a robot lacks emotion, it would be less capable of performing distinctive assessments compared to a human. The second is that of assigning responsibility for the actions of a robot in warfare. Culpability for the perpetration of war crimes falls on clearly identifiable parties when human beings act wrongly, but presents major issues when the perpetrator is not human.

The United States is not taking responsible action to avert these problems. The only policy preventing the employment of LAWS by the U.S. is a Department of Defense (DoD) Directive that U.S. autonomous robots should not target humans.[iv] There is no legal imperative stopping the DoD from revoking or failing to renew this policy, and the DoD has labeled increasing autonomy in unmanned systems a "high priority" that will be "critical to future conflicts."[v] The U.S. Delegation to the 2016 Convention on Certain Conventional Weapons recently clarified, "[DoD Directive 3000.09] does not establish a U.S. position on the future development of LAWS – it neither encourages nor prohibits the development of future systems."[vi] In the context of this international discussion, the U.S. should seize the opportunity to affirm its position as a champion of human rights both domestically and internationally. The U.S. Senate Armed Services Committee should lead an international ban on LAWS by implementing a standard throughout military operations that any act that will directly result in loss of life requires active confirmation from a human being at a minimum.

III. Distinction

LAWS can be violently objective, "[evaluating] information much faster and from more sources than human soldiers before responding with lethal force."[vii] However, they have a shortcoming that no amount of speed and no increase in signal volume can overcome: they lack emotion. The ability to contextualize situations based on emotion and to empathize based on how humans understand interpersonal interactions is both a distinctly human trait and critical to decision-making. Research shows that rather than serving as an impediment to good decision-making, emotion may be essential in shaping the decisions that we make and allowing us to consider factors beyond those that we can consciously examine.[viii] Variables that humans struggle to quantify, ingrained in our brains after millions of years of evolution and developed over lifetimes of human contact, come into play when a person decides to kill someone else, and will not when a robot does so.

For example, Bonnie Docherty proposes a scenario that illuminates the basic human ability to contextualize a situation that a robot would struggle to perceive. Imagine that a child approaches a soldier, and his mother, frightened by the soldier's gun, runs screaming towards the child to pull him away. A human, whose empathy should help them understand the context of the situation, should assume prudent risk, refraining from violence. A robot would recognize a frenetic person running towards a friendly force and might fire.[ix]

Emotion also plays a critical role in the exercise of morality, with research showing that emotion is the impetus behind moral appraisals. Emotional responses developed around three years of age allow humans to distinguish between moral norms like "it is wrong to pull a classmates hair" and conventional norms like "it is wrong to chew gum in class."[x] Although computers may mimic human decision-making, they can never understand this distinction; their lack of emotions means they cannot recognize moral questions and decide upon a moral course of action. Formulations of the law of armed conflict regularly seek to ensure that people are not treated as mere means to an end, indicated in part by the recurrence of the doctrine of double effect to justify collateral damage in different systems throughout time.[xi] In contrast to the historic significance of this precept, an algorithm can never treat a person as a moral object, only as a variable. The distinction involved in the act of killing inherently entails moral considerations, and should not be exercised by devices that cannot make a moral decision.

Some argue that LAWS's lack of emotions will serve a humanitarian end. Ronald Bailey points out that emotions can drive soldiers to do terrible things in the name of vengeance or fear, perhaps leading to an atrocity.[xii] Gregory and Diana Noone expand on this premise, pointing out the potential that LAWS might "reduce human casualties, collateral damage, and war crimes by making war less inhumane through lessening the human element from warfare."[xiii]

While it is true that in situations like Mai Lai or the Rwandan Genocide, emotions have aroused the vilest of human dispositions, it is hard to quantify how many times throughout history they have restrained those urges. While an army of overzealous soldiers makes a dangerous foe for liberty, an army of unquestioning, unfeeling robots is an ideal tool for a repressive dictator who need not fear his troops turning on him.[xiv] Of these two scenarios, a force acting illegally based on emotion and another acting illegally based on orders, the latter has historically proved much more deadly. Even in the cases of the aforementioned atrocities that aimed to highlight the shortcomings of human emotion, a robot that blindly followed the orders of a superior would have fared no better. LAWS do not protect warfare from emotion, they simply shrink the group that can exercise it from all combatants to a few leaders, and in doing so deprive those who might exercise it to disobey illegal orders of the opportunity to do so. In the twentieth century people who served as pawns in grand killing schemes became characterized as "cogs in the machine" – they as individuals did not drive the killing, but their actions were nevertheless critical for its realization.[xv] Replacing this mechanism with a literal machine will only make it more efficient.

IV. Responsibility

If a LAWS does commit an atrocity, an unprecedented problem arises – assigning culpability. Accountability in warfare is essential to both pursue justice in the aftermath of a war crime and to act as a deterrent to prevent further war crimes. By holding entities accountable, there is an incentive to take action to prevent future occurrences.[xvi] The inability to do so with LAWS is a major shortcoming and should disqualify their use.

Several attempts to extend regulations concerning war crimes into the realm of LAWS involve the punishment of the commander who decides to employ the LAWS. These are based on existing rules that punish commanders for failure to prevent their subordinates from committing war crimes when they might have reasonably known about the committal of or the intention to commit these crimes.[xvii] These cases, however, are of recklessness or negligence employing a LAWS in a manner for which it is not intended; they do not address the tougher question of how to assign responsibility when the robot is employed as intended, but does not deliver as anticipated.

The application of strict liability for the LAWS producer, where their lack of intent or negligence would not relieve them of legal responsibility, would seem to accomplish this latter end. However, any producer that needs to pay for LAWS shortcomings under strict liability could simply externalize these costs by including them in the price of the product rather than necessarily making a better product, as occurred with the Ford Pinto in the nineteen seventies.[xviii] Additionally, attempts to apply this rule to LAWS fall victim to the "Responsibility Gap" wherein the autonomy of the robot undermines the responsibility of its producer. The nature of LAWS, analyzing human interactions and society and making judgments, entails an enormous understanding of these topics, one that cannot be hard coded but instead must be learned. The more a LAWS learns from its environment, the better it should grow at exercising discretion, but simultaneously the more autonomous it becomes, evolving from its original coding.[xix] As LAWS grow to reflect their education more than their producers, they become less a product and more an autonomous entity.

This is partly why some call for the robots themselves to be held accountable for their actions.[xx] The implication is that such an advanced robot would refrain from acting out of line when there is the threat of castigation more than when simply given the same instructions without the costs associated with being wrong; in other words, it fears punishment. Cognitive research and the psychology of of moral motivation challenge the presumption that technology could ever yield such a device, but if humanity does design intelligences that can judge social and moral situations as well as us, and have emotions themselves, employing them as enslaved war machines should not be a consideration.[xxi]

Daniel Hammond proposes that the solution is to hold the state employing the LAWS accountable for all of its actions.[xxii] Though this may deter and prevent war crimes because a state would seek to ensure that certain safety protocols are followed, it is not a thorough solution. As Kelly Cass highlights, reparations do not fulfill the need for retributive justice in many cultures, and failure to hold individuals accountable for unlawful murder could lead to violent reprisals.[xxiii] Furthermore, setting the bar for pecuniary deterrence high enough that states will take concerns seriously is virtually impossible with global economies of trillions of dollars.

V. Counterpoint: Inevitability/Progress
Historically, bans on chemical weapons, blinding lasers, and torture have been successful, while bans on crossbows and aerial bombardment have not. Rebecca Crootof explains that the determining factors of a ban's efficacy are the availability of alternatives to accomplish the same purpose and the narrow tailoring of the prohibition.[xxiv] The policy recommendation herein is narrowly focused, demanding only that a human being decide another person should die instead of leaving that decision to a mathematical formula. Additionally, it doesn't seek to ban semi-autonomous weapon systems, leaving a viable option available for those countries looking to move soldiers farther from the battlefield.

Despite this, some authors insist that LAWS are the natural progress for weapons and a ban cannot be effective.[xxv] These authors ignore the fact that self-interested nations will not only consider the offensive capabilities of these weapons, but also the potential to have to face them. It is the same principle that drives states to agree not to torture prisoners both because it is inherently wrong, and as a policy to ensure their own citizens are treated well.

In the short-term, more technologically advanced states would have an advantage through the development of LAWS. However, as with the nuclear arms race, technology will spread and long-term results will disrupt this hegemony. Concerns that the U.S. might put itself at risk for a preemptive strike are unfounded, as the lack of a global treaty would become evident well before a truly legitimate threat from LAWS, allowing the U.S. to change its policy in the future if need be. This is a convenience that will not be present if the world waits until production

has begun to consider a ban. While enforcement of a ban would be difficult, Human Rights Watch correctly points out that enforcing a voluntary set of regulations, as the U.S. has currently recommended, is even more difficult.[xxvi]

VI. The Way Forward

An international ban on LAWS is not something that the world can wait to consider – the threats they pose are present now. War and killing are violent, the capacity to use this violence cannot be granted to a computer. Action by the U.S. Senate to assure that future technologies never move the decision to kill out of the hands of a human being sets a precedent that the Executive Branch must follow, and that the world needs. In the aftermath of such a law, advocacy for a binding UN Resolution based on the regulations proposed herein and by participants in the Convention on Certain Conventional Weapons can drive a global change. The U.S. is in a position to lead this movement, and it should act to preserve a future that will be decided by humans.

References

[i] Rabiroff, Jon. "Machine Gun-toting Robots Deployed on DMZ." Stars and Stripes. July 12, 2010. Accessed July 12, 2016. http://www.stripes.com/news/pacific/korea/machine-guntoting-robots-deployed-on-dmz-1.110809.

[ii] Crootof, Rebecca. 2015. "The Killer Robots Are Here: Legal and Policy Implications." *Cardozo Law Review* 36, 1849. *LexisNexis Academic: Law Reviews*, EBSCO*host*, accessed November 3, 2015

[iii] Docherty, Bonnie Lynn. *Losing Humanity: The Case against Killer Robots*. New York, N.Y.: Human Rights Watch, (2012): 8, accessed November 3, 2015.

[iv] U.S. Department of Defense, *Autonomy in Weapon Systems*, by Ashton B. Carter. Department of Defense Directive Number 3000.09, November 21, 2012. http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf.

[v] Crootof, Rebecca. 2015. "The Killer Robots Are Here: Legal and Policy Implications." *Cardozo Law Review* 36, 1850. *LexisNexis Academic: Law Reviews*, EBSCO*host*, accessed November 3, 2015.

[vi] Convention on Certain Conventional Weapons, Informal Meeting of Experts on Lethal Autonomous Weapons Systems, *U.S. Delegation Opening Statement*, by Michael W. Meier. Geneva, Switzerland, April 11, 2016. http://www.unog.ch/80256EDD006B8954/(httpAssets)/EFF7036380934E5EC1257F920057989A/$file/2016_LAWS+MX_GeneralExchange_Statements_United+States.pdf.

[vii] Bailey, Ronald. 2015. "Let slip the robots of war: lethal autonomous weapon systems might be more moral than human soldiers." *Reason Magazine*, 2015. 16. *Academic OneFile*, EBSCO*host*, accessed December 6, 2015.

[viii] Anderson, Christopher J. "The Functions of Emotion in Decision Making and Decision Avoidance." In Do Emotions Help or Hurt Decisionmaking?: A Hedgefoxian Perspective, edited by Vohs Kathleen D., Baumeister Roy F., and Loewenstein George, 183-202. Russell Sage Foundation, 2007. http://0-www.jstor.org.usmalibrary.usma.edu/stable/10.7758/9781610445436.12.

ix Docherty, Bonnie Lynn. *Losing Humanity: The Case against Killer Robots*. New York, N.Y.: Human Rights Watch, (2012): 31, accessed November 3, 2015.

x Tiberius, Valeria. "Emotion and Moral Judgment." In *Moral Psychology*, 69-86. New York: Routledge, 2015.

xi Coleman, Stephen. "Ethical Theory - An Overview." In *Military Ethics: An Introduction with Case Studies*, 8-31. New York: Oxford University Press, 2013.

xii Bailey, Ronald. 2015. "Let slip the robots of war: lethal autonomous weapon systems might be more moral than human soldiers." *Reason Magazine*, 2015. 16. *Academic OneFile*, EBSCO*host*, accessed December 6, 2015.

xiii Noone, Gregory P., and Diana C. Noone. 2015. "THE DEBATE OVER AUTONOMOUS WEAPONS SYSTEMS." *Case Western Reserve Journal Of International Law* 47, 29. *Business Source Premier*, EBSCO*host*, accessed November 3, 2015.

xiv Docherty, Bonnie Lynn. *Losing Humanity: The Case against Killer Robots*. New York, N.Y.: Human Rights Watch, (2012): 4, accessed November 3, 2015.

xv Ohlin, Jens, "The Combatant's Stance: Autonomous Weapons on the Battlefield." *International Law Studies* 92, (2016): 1-30, http://stockton.usnwc.edu/cgi/viewcontent.cgi?article=1440&context=ils.

xvi Hammond, Daniel N. 2015. "Autonomous Weapons and the Problem of State Accountability." *Chicago Journal Of International Law* 15, no. 2: 662. *International Security & Counter Terrorism Reference Center*, EBSCO*host*, accessed November 3, 2015.

xvii Schmitt, Michael N. and Thurnher, Jeffrey S. "'Out of the Loop': Autonomous Weapon Systems and the Law of Armed Conflict." *Harvard National Security Journal 231*. 2013. Available at SSRN: http://ssrn.com/abstract=2212188.

xviii Leggett, Christopher. "THE FORD PINTO CASE: THE VALUATION OF LIFE AS IT APPLIES TO THE NEGLIGENCE-EFFICIENCY ARGUMENT." Wake Forest University. 1999. Accessed July 13, 2016. http://users.wfu.edu/palmitar/Law&Valuation/Papers/1999/Leggett-pinto.

xix Johnson, Deborah G. "Technology with No Human Responsibility?" *Journal of Business Ethics* 127, no. 4 (March 2014): 708. Accessed July 13, 2016. doi:10.1007/s10551-014-2180-1.

xx Johnson, Deborah G. "Technology with No Human Responsibility?" *Journal of Business Ethics* 127, no. 4 (March 2014): 710. Accessed July 13, 2016. doi:10.1007/s10551-014-2180-1.

xxi Song, Daegene. "Non-Computability of Consciousness." *NeuroQuantology* 5, no. 4 (2007). doi:10.14704/nq.2007.5.4.146.

xxii Hammond, Daniel N. 2015. "Autonomous Weapons and the Problem of State Accountability." *Chicago Journal Of International Law* 15, no. 2: 655-656. *International Security & Counter Terrorism Reference Center*, EBSCO*host*, accessed November 3, 2015.

xxiii Cass, Kelly. "Autonomous Weapons and Accountability: Seeking Solutions." *Loyola Law Review* 48, no. 3 (2015): 1017-70. Accessed July 13, 2016. http://digitalcommons.lmu.edu/llr/vol48/iss3/11/.

xxiv Crootof, Rebecca. 2015. "The Killer Robots Are Here: Legal and Policy Implications." *Cardozo Law Review* 36, 1857. *LexisNexis Academic: Law Reviews*, EBSCO*host*, accessed November 3, 2015.

xxv Lewis, John. 2015. "The case for regulating fully autonomous weapons." *Yale Law Journal* no. 4: 1316. *Academic OneFile*, EBSCO*host*, accessed November 3, 2015.

xxvi Human Rights Watch. "Fully Autonomous Weapons: Questions and Answers." Presented at the Harvard Law School's International Human Rights Clinic, October, 2013. https://www.hrw.org/sites/default/files/supporting_resources/10.2013_killer_robots_qa.pdf.

Author Bio – Brian Fydenkevez is a recent graduate of the United States Military Academy at West Point, where he majored in Physics. He is a Second Lieutenant in the U.S. Army and is currently studying to be an Ordnance officer at the U.S. Army Logistics University in Fort Lee, VA.