

College of Saint Benedict and Saint John's University

DigitalCommons@CSB/SJU

---

Computer Science Faculty Publications

Computer Science

---

1-3-2022

## Can lethal autonomous weapons be just?

Noreen L. Herzfeld

College of Saint Benedict/Saint John's University, [nherzfeld@csbsju.edu](mailto:nherzfeld@csbsju.edu)

Robert H. Latiff

Follow this and additional works at: [https://digitalcommons.csbsju.edu/csci\\_pubs](https://digitalcommons.csbsju.edu/csci_pubs)



Part of the [Applied Ethics Commons](#), [Artificial Intelligence and Robotics Commons](#), and the [Military and Veterans Studies Commons](#)

---

### Recommended Citation

Herzfeld N, Latiff R. 2022. Can lethal autonomous weapons be just? *Peace Review* 33(2): 213-219. <https://doi.org/10.1080/10402659.2021.1998750>

This is an Accepted Manuscript version of the following article, accepted for publication in *Peace Review*: Herzfeld N, Latiff R. 2022. Can lethal autonomous weapons be just? *Peace Review* 33(2): 213-219. <https://doi.org/10.1080/10402659.2021.1998750>

It is deposited under the terms of the [Creative Commons Attribution-NonCommercial License](#), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

# Can Lethal Autonomous Weapons Be Just?

Noreen Herzfeld and Robert H. Latiff

## QUERY SHEET

This page lists questions we have about your paper. The numbers displayed at left are hyperlinked to the location of the query in your paper.

The title and author names are listed on this sheet as they will be published, both on your paper and on the Table of Contents. Please review and ensure the information is correct and advise us if any changes need to be made. In addition, please review your paper as a whole for typographical and essential corrections.

Your PDF proof has been enabled so that you can comment on the proof directly using Adobe Acrobat. For further information on marking corrections using Acrobat, please visit <http://journalauthors.tandf.co.uk/production/acrobat.asp>; <https://authorservices.taylorandfrancis.com/how-to-correct-proofs-with-adobe/>

The CrossRef database ([www.crossref.org/](http://www.crossref.org/)) has been used to validate the references.

## AUTHOR QUERIES

- Q1** Please provide the missing closing double quote for sentence “lives being put at risk, significantly ...”
- Q2** Please provide the missing closing double quote for sentence “a president who sends someone’s ...”
- Q3** Please provide the short biography for author name “Robert H. Latiff.”



# Can Lethal Autonomous Weapons Be Just?

NOREEN HERZFELD AND ROBERT H. LATIFF

In 2018 the United States Department of Defense (DoD) created a new Joint Artificial Intelligence Center to study the adoption of AI by the military. Their strategy, outlined in a document entitled, “Harnessing AI to Advance Our Security and Prosperity,” proposes to accelerate the adoption of AI in the military by fostering a culture of experimentation and calculated risk taking, noting that AI will change the character of the future battlefield and, even more, the pace of battle. Is there any way to ensure that this future battlefield will be just? Can the age-old precepts of just warfare help guide our militaries as we develop and deploy autonomous weapons?

War kills people. It destroys the environment. It replaces trust with fear and strips millions of their rights, livelihood, and home. When can one justify these as responsible or ethical acts? Just war theory has, over the centuries, provided an amorphous set of rules and principles, rooted in Christian thought and scripture, both for when to wage war and for acceptable conduct within a war. Warfare, and the means of warfare, constantly change. Just as the advent of nuclear weapons forced twentieth-century theologians to reevaluate the justness of war, the advent of lethal autonomous weapons calls twenty-first century theologians to do likewise. Can we ensure ethical and moral responsibility in an era when machines make the decisions?

While rules of conduct in war can be found in Old Testament texts, the pre-Christian thought of Aristotle, and the works of Augustine, it is Thomas Aquinas, in his *Summa Theologicae*, who first lays out the general outline of what we now regard as just war theory. These principles were adjusted and universalized by later scholastics and further rethought by twentieth century theologians and philosophers, due to the advent of nuclear weapons. The coming use of AI and LAWS as tools of war once again require us to rethink our justifications for initiating and executing hostilities.

41 AI requires us to consider what difference autonomy makes in the  
42 deployment of weaponry. While bombs, land mines, missiles, and drones do  
43 not always involve a direct human decision as to when or where they wreak  
44 their havoc, these weapons do not make decisions. They cannot decide not to  
45 explode when triggered. Nor can they choose a specific target on their own.  
46 A landmine targets whoever steps on it; a bomb maims or kills whoever is  
47 within range. Artificial intelligent weapons inaugurate a new era in weaponry  
48 that differs in kind rather than merely by degree. It becomes essential then to  
49 reconsider how the concepts of just war theory are to be applied.

50 The precepts of just warfare are traditionally broken into two sets:  
51 *jus ad bellum*, when it is justifiable to go to war, and *jus in bello*, how to  
52 justly conduct combat. *Jus ad Bellum*, when to go to war, includes the fol-  
53 lowing guidelines: First, the cause must be just. One can fight for many  
54 reasons but few are grave enough to justify the loss of life and property  
55 war engenders, for, as Pope Pius XII stated on the eve of World War II,  
56 “Nothing is to be lost with peace; everything can be lost with war.”  
57 Traditionally, force may be used to resist attack, protect innocent life in  
58 imminent danger, and, as phrased by the U.S. Catholic bishops, to  
59 “correct a grave, public evil” such as genocide or a massive violation of a  
60 group’s basic human rights. War must be declared by a legitimate authority.  
61 It must be the last resort after all peaceful alternatives have been seriously  
62 tried and exhausted. War should be fought only for the purpose that initiated  
63 it, to correct the suffering or grave evil that was its cause, never out of mal-  
64 ice, ethnic hatred, sheer love of fighting, as a distraction from domestic diffi-  
65 culties, or for economic or material gain. The anticipated benefits of waging  
66 war must outweigh the expected harm, and, finally, there must be a reason-  
67 able chance of success.

68 The guidelines for *Jus in Bello*, or right conduct while fighting, trad-  
69 itionally include discrimination, proportionality, and right intention. Acts  
70 of war should discriminate between enemy combatants and noncomb-  
71 atants and civilians should be protected as far as possible. No more force  
72 should be used than what is necessary to attain one’s ends. The waging of  
73 war must be solely with the intention of righting the wrong and achieving  
74 a legitimate military objective. Acts of vengeance and indiscriminate vio-  
75 lence are forbidden. The use of weapons or methods that are intrinsically  
76 evil, such as mass rape, forcing enemy combatants to fight against or  
77 betray their own side, or weapons whose effects are uncontrollable, such  
78 as chemical or biological weapons, is forbidden.

80 **H**ow do autonomous weapons fit with these principles? LAWS exist  
81 along a scale of autonomy. At one end we have traditional “fire and  
82 forget” weapons, such as guided missiles in which a human operator  
83

84 selects a target and launches the missile that then uses sensors and algo-  
85 rithms to complete the task. These clearly depend on a human operator  
86 being “in the loop.” “Human on the loop” are weapons that act under  
87 human supervision, where an operator has the ability to monitor the  
88 weapon and halt or alter its engagement. They are semi-autonomous.  
89 Fully autonomous weapons, where the human is “out of the loop” can  
90 select and engage targets without any intervention by a human operator.  
91 These systems often depend on machine learning and, by definition, one can-  
92 not predict with certainty which targets it will attack or why they were  
93 chosen. The advent of flight inaugurated a new era of warfare, releasing  
94 armies from physical presence on the field of battle. LAWS will inaugurate  
95 a third era, releasing soldiers from the mental decisions of the battlefield as  
96 well. Their usage represents the crossing of a new “moral Rubicon.”

97 On a practical level, LAWS will not only remove too many of the  
98 psychological barriers to war, but will also privilege offense over defense.  
99 Thus, their mere availability could make it difficult to say that war is a  
100 last resort. War becomes a too-easy, attractive alternative. The use of  
101 LAWS removes the constraint of soldiers “lives being put at risk, signifi-  
102 cantly lowering the cost of an attack. Thus, the threshold of entry into a  
103 war could be substantially lowered. As an editorial in *The Economist*  
104 points out, “a president who sends someone’s son or daughter into battle  
105 has to justify it publicly, as does the congress responsible for appropri-  
106 ations and a declaration of war. But if no one has children in danger, is it  
107 a war?”

108 Examples of LAWS in use or development show multiple options in  
109 autonomy. The HAROP loitering missile can either be controlled via a  
110 two-way data link for “human-in-the-loop” operation or programmed to  
111 autonomously recognize and attack high-value targets. The EGIS naval  
112 air defense system, used by the navies of the United States, Australia,  
113 Japan, Norway, Republic of Korea and Spain, is able to search both in the  
114 air and on the surface and track and guide missiles, deciding autono-  
115 mously when and where to fire. It can function fully autonomously or in  
116 “human on the loop” mode with operators having the option to override  
117 its decisions.

118 These and other weapons under development present military  
119 commanders with a variety of incentives for use. They can process vast  
120 amounts of data and operate at speeds and levels of precision far beyond  
121 human capabilities, including making rapid decisions to changing circum-  
122 stances. They can operate in harsh and difficult environments. They are  
123 less expensive than human troops and can work long hours without tiring.  
124 They can carry out orders with fewer mistakes. Most important, they keep  
125 soldiers out of physically and psychologically dangerous or deadly  
126

127 environments. These advantages, however, do not come without costs. In  
128 what ways does the advent of these weapons affect our decisions on when  
129 to wage war, how to wage war, and who is responsible for the acts of  
130 war? We turn to an examination of one consideration from each category  
131 of just war theory to provide a brief and partial answer to this question.  
132

133 **W**hile much has been written about autonomous weapons themselves,  
134 justice in going to war is not only a question of the availability of  
135 autonomous weapons, but also of autonomous decision support systems.  
136 For example, the U.S. military is racing to incorporate AI into not only  
137 individual weapons, but also into higher-level command and control systems.  
138 The DoD's Joint All Domain Command and Control concept aims  
139 to centralize planning and execution of its operations, including space and  
140 cyber. Soon, AI will fuze data from worldwide sensors to create a single  
141 "common operating picture" for decision makers.

142 The military services have a number of related programs that are  
143 designed to demonstrate such capabilities. The Army's Project  
144 Convergence and the Air Force's Advanced Battle Management System  
145 incorporate AI to determine the best pairing between shooters and targets.  
146 Similarly, the Defense Advanced Research Project Agency's Mosaic  
147 Warfare program seeks to employ AI to network systems and sensors, pri-  
148 oritize sensor data, and autonomously determine the optimal composition  
149 of forces. As AI systems mature, algorithms will provide commanders  
150 with viable courses of action based on real-time analysis of the battle-  
151 space, thereby increasing the speed of decision-making. One ongoing  
152 DARPA program would enable the system to autonomously observe the  
153 situation, orient to what is observed, decide the best course of action,  
154 "and then act."

155 The worry here is in the bias that such a support system might con-  
156 tain. If the bias is, naturally, toward our side's dominance and ultimate  
157 victory, how do we ensure that the decisions by the command and control  
158 system are not themselves biased in that direction? Can we be certain that  
159 the system has looked at every possibility short of war and given a fair  
160 judgment? Even senior military and intelligence officials have expressed  
161 concerns about bias in algorithms developed for national security pur-  
162 poses. If AI controlled systems are calling the shots, concepts of just  
163 cause, legitimate authority, or right intention lose their meaning, since  
164 machines are incapable of such morality-based decisions.

165 Once at war, the principles of *jus in bello* demand that one act with  
166 restraint, refraining from gratuitous killing of civilians, from excess  
167 destruction, and that soldiers conduct themselves with virtue and propri-  
168 ety. Robotist Ron Arkin has argued that LAWS have the potential to act  
169

170 more virtuous than humans. Arkin cites a report from the Surgeon  
171 General's Office assessing the battlefield ethics of U.S. soldiers and  
172 marines in which ten percent reported mistreating noncombatants and  
173 roughly thirty percent reported facing ethical situations to which they did  
174 not know how to respond. Soldiers, under pressure, often react emotion-  
175 ally, out of fear or anger.

176 An analysis of civilian casualties in the second Iraqi war found that  
177 most were either the result of ethnic cleansing or caused by indiscriminate  
178 fire between sides. According to Arkin, robots could be programmed  
179 without emotions, so would never react out of panic or vengeance.  
180 Similarly, since they are not mortal, as we are, they would not act out of  
181 fear or a need for self-protection. They would follow orders more exactly  
182 and could integrate information regarding a changing battle scenario faster  
183 before responding with lethal force, thus acting with more precision and  
184 fewer mistakes. Arkin believes AIs could better discriminate between  
185 combatants and noncombatants, thus committing fewer war crimes and  
186 reducing civilian casualties.

187 We do not agree. Were the laws of war reducible to a set of simple  
188 rules you would think we would have found that set by now. Morality is  
189 inherently both ambiguous and context sensitive. It is also difficult to  
190 instantiate ~~a general rule~~. For example, programming a robot to discrimi-  
191 nate between a combatant and a civilian using facial recognition might be  
192 easy enough in the case of an individual assassination but remarkably dif-  
193 ficult in the general context of a counterinsurgency. Arkin is correct in  
194 noting that soldiers often violate the principle of right intention, acting  
195 out of fear or anger. But they also act out of altruism and mercy. The con-  
196 science of a human soldier can act as a check on unjustifiable commands  
197 or illegal orders. AIs, as they currently stand, have no intrinsic intentions.  
198 While LAWS may make decisions without direct human control, so far,  
199 they cannot reason about those decisions. Thus, there remains a direct  
200 causal chain between the machine's behavior and its programmers. A  
201 machine with true agency would have a further ability to reason independ-  
202 ently about its own actions and unpredictably change course, should it  
203 consider those actions unethical or in violation of an overarching value or  
204 intention. Yet, where would these intentions come from?

206  
207 **W**e learn our social responsibilities through a lifetime of experience,  
208 gradually, from our parents, our peers, our faith traditions, and, for  
209 a soldier, from his or her fellow soldiers, commanding officers, and basic  
210 training. Most soldiers report that the greatest motivating factor for their  
211 actions on the battlefield is their sense of solidarity with and responsibility  
212 for their fellow soldiers. The so-called "military ethos" is hugely



213 important. Soldier behavior on the battlefield is heavily weighted with  
214 emotions—such as courage, loyalty to mates, altruism, empathy, fear,  
215 guilt, etc. As yet, we cannot instill these into a machine. The presence of  
216 these emotions in humans and their absence in machines could create a  
217 difficult or impossible dynamic in battlefield unit cohesion.

218 Expecting LAWS to follow the laws of just warfare assumes not  
219 only that these percepts are codifiable or learnable by an AI but that they  
220 are what would actually be programmed. An AI's programming might be  
221 altered by a bad actor. Current AI systems are notoriously subject to hack-  
222 ing and corrupted data. Should autonomous weapons become "small,  
223 smart, cheap, and abundant," it will be hard to keep them out of the hands  
224 of terrorists and non-state actors who would not have the same interest in  
225 following the rules of warfare. Even lacking these scenarios, however,  
226 how likely is a military to prioritize ethics over victory? We fear that a  
227 nod to ethical principles could easily degenerate into mere "window  
228 dressing" for the public while the true goal programmed into LAWS  
229 would be to win at all costs.

230 Sometimes moral behavior means breaking the rules. Most of the  
231 disputes between Jesus and the Pharisees recorded in the Gospels hinged  
232 precisely on Jesus or his disciples breaking a rule or religious convention.  
233 The spirit of the law does not always match the letter. In the Iraq war, the  
234 Mahdi Militia used a child as a forward observer. U.S. forces did not  
235 shoot the child even though the conventions of war would allow this.  
236 Could an AI be programmed with sufficient nuance to make this judgment  
237 call?

238 Under military law, a commanding officer is held responsible for the  
239 actions of those under his or her command if those actions could in any  
240 way have been foreseen or prevented. Would LAWS be sufficiently pre-  
241 dictable so that commanders would have enough assurance to risk using  
242 them? If a weapon is unpredictable, can the commander justly deploy it?  
243 The rules of war do not specify the role of human judgment. As we  
244 design and build AIs to aid us in our tasks, there are two directions we  
245 can take. The first is mimesis, designing machines to take our place. The  
246 second is symbiosis, leveraging the distinctive strengths of the computer  
247 to work together with human beings.

248 Perhaps the greatest talent computers bring to warfare is speed. But  
249 this asset can also be a liability. While most commanders express a desire  
250 for autonomous weapons to have humans in or at least on the loop, how  
251 much control can they actually have if decisions in the field are made at a  
252 speed that humans are unable to follow? The tempo of war has steadily  
253 accelerated, increasing dramatically in recent years. At what point, as the  
254 speed of warfare exceeds human capacities, might we be forced to cede  
255



all decision making to the machines? As AI moves from tactical to strategic decisions this could eviscerate any meaning from the concept of “mission command.” As computers gain autonomy, we risk losing it.

At a recent workshop sponsored by the law faculty at Penn State, military commanders, both active and retired, expressed their personal distaste for LAWS. We join with these commanders in hoping that we never reach the point where LAWS outstrip human commanders’ ability to control them. In all but the most servile applications, computer-human symbiosis is preferable to full autonomy. Whether on the field of battle or in the workplace, human dignity depends on our working with our tools rather than letting them supplant us and this is at its most important in matters that involve questions of life and death. We can only hope that the long tradition of just warfare, a spirit of humility, and, as Pope Francis counsels us in his recent encyclical, *Fratelli Tutti*, a spirit of acknowledging the fraternity and sorority of all humans will keep military commanders asking not just whether a certain course is expedient but whether it is just, decent, and moral.

This essay is part of project ARRS J6-1813: *Creations, Humans, Robots: Creation Theology Between Humanism and Posthumanism*, funded by the Slovenian Research Agency.

## RECOMMENDED READINGS

- Arkin, Ronald. 2015. “The Case for Banning Killer Robots: Counterpoint.” *Communications of the ACM* 58 (12):46–47.
- Crotoof, Rebeca. 2016. “War Torts: Accountability for Autonomous Weapons.” *University of Pennsylvania Law Review* 164 (6):1347–1403.
- Johnson, James T. 2011. *Ethics and the Use of Force*. London: Routledge.
- Payne, Kenneth. 2018. *Strategy, Evolution, and War: From Apes to Artificial Intelligence*. Washington, DC: Georgetown University Press.
- Scharre, Paul. 2018. *Army of None: Autonomous Weapons and the Future of War*. New York: Norton.
- Solis, Gary. 2016. *The Law of Armed Conflict*. New York: Cambridge University Press.
- Walzer, Michael. 1977. *Just and Unjust Wars: A Moral Argument with Historical Illustrations*. New York: Basic Books.

---

Noreen Herzfeld is the Nicholas and Bernice Reuter Professor of Science and Religion at St. John’s University and the College of St. Benedict and a research associate with the Science and Research Center, Koper Slovenia. Herzfeld is the author of: *In Our Image: Artificial Intelligence and the Human Spirit* (Fortress, 2002), *Technology and Religion: Remaining Human in a Co-Created World* (Templeton, 2009), *The Limits of Perfection in Technology, Religion, and Science* (Pandora, 2010) and editor of *Religion and the New Technologies* (MDPI, 2017). E-mail: [nherzfeld@csbsju.edu](mailto:nherzfeld@csbsju.edu)

Q3 ~~Maj Gen (Ret) Robert H. Latiff~~, E-mail: [rlatiff@nd.edu](mailto:rlatiff@nd.edu)