

DOI:10.1145/2835965

Counterpoint: Ronald Arkin

LET ME UNEQUIVOCALLY state: *The status quo with respect to innocent civilian casualties is utterly and wholly unacceptable.* I am not Pro Lethal Autonomous Weapon Systems (LAWS), nor for lethal weapons of any sort. I would hope that LAWS would never need to be used, as I am against killing in all its manifold forms. But if humanity persists in entering into warfare, which is an unfortunate underlying assumption, we must protect the innocent noncombatants in the battlespace far better than we currently do. Technology can, must, and should be used toward that end. Is it not our responsibility as scientists to look for effective ways to reduce man's inhumanity to man through technology? Research in ethical military robotics could and should be applied toward achieving this goal.

I have studied ethology (animal behavior in their natural environment) as a basis for robotics for my entire career, spanning frogs, insects, dogs, birds, wolves, and human companions. Nowhere has it been more depressing than to study human behavior in the battlefield (for example, the Surgeon General's Office 2006 report¹⁰ and *Killing Civilians: Method, Madness, and Morality in War*.⁹). The commonplace occurrence of slaughtering civilians in conflict over millennia gives rise to my pessimism in reforming human behavior yet provides optimism for robots being able to exceed human moral performance in similar circumstances. The regular commission of atrocities is well documented both historically and in the present day, reported almost on a daily basis. Due to this unfortunate low bar, my claim that robots may be able to eventually outperform humans with respect to adherence to international humanitarian law (IHL) in warfare (that is, be more humane) is credible. I have the utmost respect for our young men and women in the battlespace, but they are placed into situations where no human has ever been designed to function. This is exacer-

bated by the tempo at which modern warfare is conducted. Expecting widespread compliance with IHL given this pace and resultant stress seems unreasonable and perhaps unattainable by flesh and blood warfighters.

I believe judicious design and use of LAWS can lead to the potential saving of noncombatant life. If properly developed and deployed it can and should be used toward achieving that end. It should not be simply about winning wars. We must locate this humanitarian technology at the point where war crimes, carelessness, and fatal human error lead to noncombatant deaths. It is not my belief that an unmanned system will ever be able to be perfectly ethical in the battlefield, but I am convinced they can ultimately perform more ethically than human soldiers.

I have stated that I am not averse to a ban should we be unable to achieve the goal of reducing noncombatant casualties, but for now we are better served by a moratorium at least until we can agree upon definitions regarding what we are regulating, and it is indeed determined whether we can realize humanitarian benefits through the use of this technology. A preemptive ban ignores the moral imperative to use technology to reduce the persistent atrocities and mistakes that human warfighters make. It is at the very least premature. History indicates that technology can be used toward these goals.⁴ Regulate LAWS usage instead of prohibiting them entirely.⁶ Consider restrictions in well-defined circum-

Is it not our responsibility as scientists to look for effective ways to reduce man's inhumanity to man through technology?

stances rather than an outright ban and stigmatization of the weapon systems. Do not make decisions based on unfounded fears—remove pathos and hype and focus on the real technical, legal, ethical, and moral implications.

In the future autonomous robots may be able to outperform humans from an ethical perspective under battlefield conditions for numerous reasons:

- ▶ Their ability to act conservatively, as they do not need to protect themselves in cases of low certainty of target identification.

- ▶ The eventual development and use of a broad range of robotic sensors better equipped for battlefield observations than humans currently possess.

- ▶ They can be designed without emotions that cloud their judgment or result in anger and frustration with ongoing battlefield events.

- ▶ Avoidance of the human psychological problem of “scenario fulfillment” is possible, a factor contributing to the downing of an Iranian Airliner by the USS *Vincennes* in 1988.⁷

- ▶ They can integrate more information from more sources far faster than a human possibly could in real time before responding with lethal force.

- ▶ When working in a team of combined human soldiers and autonomous systems, they have the potential of independently and objectively monitoring ethical behavior in the battlefield by all parties and reporting infractions that might be observed.

LAWS should not be considered an end-all military solution—far from it. Limited circumstances for their use must be utilized. Current thinking recommends:

- ▶ Specialized missions only where bounded morality,^{a,1} applies, for example, room clearing, countersniper operations, or perimeter protection in the DMZ.^b

- ▶ High-intensity interstate warfare,

a Bounded morality refers to adhering to moral standards within the situations that a system has been designed for, in this case specific battlefield missions and not in a more general sense.

b For more specifics on these missions see Arkin, R.C., *Governing Lethal Behavior in Autonomous Systems*, Chapman-Hall, 2009.

not counterinsurgencies, to minimize likelihood of civilian encounter.

► Alongside soldiers, not as a replacement. A human presence in the battlefield should be maintained.

Smart autonomous weapon systems may enhance the survival of noncombatants. Consider Human Rights Watch's position on the use of precision-guided munitions in urban settings—a moral imperative. LAWS in effect may be mobile precision-guided munitions resulting in a similar moral imperative for their use. Consider not just the possibility of LAWS making a decision when to fire, but rather deciding when *not* to fire (for example, smarter context-sensitive cruise missiles). Design them with runtime human overrides to ensure meaningful human control,¹¹ something everyone wants. Additionally, LAWS can use fundamentally different tactics, assuming far more risk on behalf of noncombatants than human warfighters are capable of, to assess hostility and hostile intent, while assuming a “First do no harm” rather than “Shoot first and ask questions later” stance.

To build such systems is not a short-term goal but will require a mid- to long-term research agenda addressing the many very challenging research questions. By exploiting bounded morality within a narrow mission context, however, I would contend that the goal of achieving better performance with respect to preserving noncombatant life is achievable and warrants a robust research agenda on humanitarian grounds. Other researchers have begun related work on at least four continents. Nonetheless, there remain many daunting research questions regarding lethality and autonomy yet to be resolved. Discussions regarding regulation of LAWS must be based on reason and not fear. Some contend that existing IHL may be adequate to afford adequate protection to noncombatants from the potential misuse of LAWS.² A moratorium is more appropriate at this time than a ban, until these questions are resolved and only then can careful, graded introduction of the technology into the battlespace be ensured. Proactive management of these issues is necessary. Other technological approaches are of course welcome, perhaps such as the creation

I say to my fellow researchers, if your research is of any value, someone somewhere someday will put it to work in a military system.

of ethical advisory systems for human warfighters to assist in their decision-making when in conflict.

Restating my main point: *The status quo is unacceptable with respect to non-combatant deaths.* It may be possible to save noncombatant lives through the use of this technology—if done correctly—and these efforts should not be prematurely terminated by a preemptive ban.

Quoting from a recent *NewsWeek* article³: “But autonomous weapon systems would not necessarily be like those crude weapons [poison gas, landmines, cluster bombs]; they could be far more discriminating and precise in their target selection and engagement than even human soldiers. A preemptive ban risks being a tragic moral failure rather than an ethical triumph.”

Similarly from the *Wall Street Journal*⁸: “Ultimately, a ban on lethal autonomous systems, in addition to being premature, may be feckless. Better to test the limits of this technology first to see what it can and cannot deliver. Who knows? Battlefield robots might yet be a great advance for international humanitarian law.”

I say to my fellow researchers, if your research is of any value, someone somewhere someday will put it to work in a military system. You cannot be absolved from your responsibility in the creation of this new class of technology simply by refusing a particular funding source. Bill Joy argued for the relinquishment of robotics research in his *Wired* article “Why the Future Doesn't Need Us.”⁵ Perhaps it is time for some to walk away from AI if their conscience so dictates.

But I believe AI can be used to save innocent life, where humans may and

do fail. Nowhere is this more evident than on the battlefield. Until that goal can be achieved, I support a moratorium on the development and deployment of this technology. If our research community, however, firmly believes the goal of achieving better performance than a human warfighter with respect to adherence to IHL is unattainable, and states collectively that we cannot ever reach this level of exceeding human morality in narrow battlefield situations where bounded morality applies and where humans are often at their worst, then I would be moved to believe our community asserts artificial intelligence in general is unattainable. This appears to contradict those who espouse their goal of doing just that.

We must reduce civilian casualties if we are foolish enough to continue to engage in war. I believe AI researchers have a responsibility to achieve such reductions in death and damage during the conduct of warfare. We cannot simply accept the current status quo with respect to noncombatant deaths. Do not turn your back on those innocents trapped in war. It is a truly hard problem and challenge but the potential saving of human life demands such an effort by our community. ■

References

1. Allen, C., Wallach, W., and Smit, I. Why machine ethics? *IEEE Intelligent Systems* (Jul./Aug. 2006), 12–17.
2. Anderson, K. and Waxman, K. Law and ethics for autonomous weapon systems: Why a ban won't work and how the laws of war can. Stanford University, The Hoover Institution (Jean Perkins Task Force on National Security and Law Essay Series), 2013.
3. Bailey, R. Bring on the killer robots. *Newsweek*; (Feb. 1, 2015); <http://bit.ly/1K3VaYK>
4. Horowitz, M. and Scharre, P. Do killer robots save lives? *Politico Magazine* (Nov. 19, 2014).
5. Joy, B. Why the future doesn't need us. *Wired* 8, 4 (Apr. 2000).
6. Muller, V. and Simpson, T. Killer robots: Regulate, don't ban. Blavatnik School of Government Policy Memo, Oxford University, Nov. 2014.
7. Sagan, S. Rules of engagement. In *Avoiding War: Problems of Crisis Management*. A. George, Ed., Westview Press, 1991.
8. Schechter, E. In defense of killer robots. *Wall Street Journal* (July 10, 2014).
9. Slim, H., *Killing Civilians: Method, Madness, and Morality in War*. Columbia University Press, New York, 2008.
10. Surgeon General's Office, Mental Health Advisory Team (MHAT) IV Operation Iraqi Freedom 05-07, Final Report, Nov. 17, 2006.
11. U.N. *The Weaponization of Increasingly Autonomous Technologies: Considering How Meaningful Human Control Might Move the Discussion Forward*. UNIDIR Resources, Report No. 2, 2014.

Ronald Arkin (arkin@cc.gatech.edu) is a Regents' Professor and is the director of the Mobile Robot Laboratory in the College of Computing at the Georgia Institute of Technology.

Copyright held by author.