

# Ethical Robots in Warfare

RONALD C. ARKIN

Robotic system capabilities have advanced dramatically over the last several decades. We now have artificially intelligent systems and robots that are stronger than humans, that can venture places where people cannot go (such as Mars), that are smarter than people in certain cases (e.g., in chess), and so on. We are no longer truly surprised when machine artifacts outperform humans in new domains.

But the outperformance of humans by artificially intelligent systems may still come as a surprise to some. It is a thesis of my ongoing research for the U.S. Army that robots not only can be better than soldiers in conducting warfare in certain circumstances, but they also can be more humane in the battlefield than humans.

Why should this surprise us? Do we believe that human warfighters exhibit the best of humanity in battlefield situations? There is strong evidence to the contrary [1]-[4] and we have developed Laws of War to criminalize those people who behave outside of acceptable international norms. Despite these regulations, they are often cast aside in the heat of

combat, for reasons such as vengeance, anger, frustration, and the desire for victory at any cost.

Robots already have the ability to carry weapons and use lethal force under the direction of a human operator. Multiple unmanned robotic systems are already being developed or are in use that employ lethal force such as the Armed Robotic Vehicle (ARV), a component of the Future Combat System (FCS); Predator and Reaper unmanned aerial vehicles (UAVs) equipped with hellfire missiles, which have already been used in combat but under direct human supervision; and the development of an armed platform for use in the Korean Demilitarized Zone, to name only a few. These and other systems are not fully autonomous in this sense: they do not currently make decisions on their own about when, or not, to engage a target. But the pressure of an increasing battlefield tempo is forcing autonomy further and further towards the point of robots making that final, lethal decision. The time available to make the decision to shoot or not to shoot is becoming too short for remote

humans to make intelligent, informed decisions in many situations that arise in modern warfare. As that time dwindles, robots will likely be given more authority to make lethal decisions on their own.

Commercially available robots already have had emotions engineered into them, e.g., the robot dog AIBO [5], so researchers, at least to some extent, have an understanding of what affect contributes to intelligent interaction with humans. It is my contention that robots can be built that do not exhibit fear, anger, frustration, or revenge, and that ultimately (and the key word here is ultimately) behave in a more humane manner than even human beings in these harsh circumstances and severe duress. People have not evolved to function in these conditions, but robots can be engineered to function well in them.

## Robot Adherence to Laws of War

In a forthcoming book entitled *Governing Lethal Behavior in Autonomous Robots*, I make the case that autonomous armed robotic platforms may ultimately reduce noncombatant casualties

and other forms of collateral damage by their ability to better adhere to the Laws of War than most soldiers possibly can. Some of the material that follows is drawn directly from this book. Many of my colleagues writing in this special section or the *IEEE Technology and Society Magazine* argue against this thesis and bring up many significant issues that must be resolved prior to such a deployment. To summarize both sides of these arguments, first, the reasons why ethical autonomy can succeed include the following.

1) The ability to act conservatively: Robots do not need to protect themselves in cases of low certainty of target identification. Autonomous armed robotic vehicles do not need to have self-preservation as a foremost drive, if at all. They can be used in a self-sacrificing manner if needed and appropriate, without reservation by a commanding officer.

2) The eventual development and use of a broad range of robotic sensors better equipped for battlefield observations than human sensory abilities.

3) Robots can be designed without emotions that cloud their judgment or result in anger and frustration with ongoing battlefield events. In addition, “Fear and hysteria are always latent in combat, often real, and they press us toward fearful measures and criminal behavior” [6]. Autonomous agents need not suffer similarly.

4) Avoidance of the human psychological problem of “scenario fulfillment” is possible, a factor believed partly contributing to the downing of an Iranian Airliner by the USS Vincennes in 1988 [7]. This phenomena leads to distortion or neglect of contradictory information in stressful

situations, where humans use new incoming information in ways that fit only their pre-existing belief patterns, a form of premature cognitive closure. Robots need not be vulnerable to such patterns of behavior.

4) Robots can integrate more information from more sources far more quickly before responding with lethal force than a human can in real-time. This information and data can arise from multiple remote sensors and intelligence (including human) sources, as part of the Army’s network-centric warfare concept [8] and the concurrent development of the Global Information Grid [9]. “Military systems (including weapons) now on the horizon will be too fast, too small, too numerous, and will create an environment too complex for humans to direct” [10].

5) When working on a team of combined human soldiers and autonomous systems as an organic asset, robots have the potential capability of independently and objectively monitoring ethical behavior in the battlefield by all parties and reporting infractions that might be observed. This presence alone might possibly lead to a reduction in human ethical infractions.

### **Additional Battlefield Robot Benefits**

Aside from these ethical considerations, autonomous robotic systems offer numerous potential operational benefits to the military: faster, cheaper, better mission accomplishment; longer range, greater persistence, longer endurance, higher precision; faster target engagement; and immunity to chemical and biological weapons, among other benefits [11]. All of these can enhance mission effectiveness and serve

as drivers for the ongoing deployment of these systems.

But this new research focuses on enhancing ethical benefits by using these systems, ideally without eroding mission performance when compared to human warfighters.

### **Arguments Against Use of Wartime Robots**

The counterarguments against the use of lethal autonomous systems are numerous as well:

- Establishing responsibility – who’s to blame if things go wrong with an autonomous robot?
- The threshold of entry into warfare may be lowered as we will now be risking machines and fewer human soldiers – this could violate the Jus ad Bellum conditions of just warfare.
- The possibility of unilateral risk-free warfare, which could be viewed as potentially unjust.
- It simply can’t be done right – it’s just too hard for machines to discriminate targets.
- The effect on military squad cohesion and its impact on the fighting force – human warfighters may not accept ethical robots monitoring their performance.
- Robots running amok – the classic science fiction nightmare.
- A robot refusing an order – the question of whether ultimate authority should vest in humans.
- The issues of overrides placed in the hands of immoral, irresponsible, or reckless individuals.
- The co-opting of an ethical robot research effort by the military to serve to justify other political agendas.
- The difficulty in winning the hearts and minds of the civilians affected by warfare if robots are allowed to kill.
- Proliferation of the technology to other nations and terrorists.

I am confident that these contrarian issues are raised in

more detail in the articles by the other authors of this magazine's special section and I will not elaborate on them here. Some are more easily dismissed than others, some are not unique to autonomous robot battlefield technology, some can be addressed by recognizing that we're dealing with bounded morality for very narrow tactical situations and are not replacing a human soldier one-for-one, and some can be addressed by suitable system design which may be long range but nonetheless feasible. Space, however, prevents a full and fair treatment of these concerns here. The goal of my research on ethical autonomous systems capable of lethal action is to provide robots with an ethical code that has been already established by humanity as encoded in the Laws of War and the Rules of Engagement. Robots must be constrained to adhere to the same laws as humans or they should not be permitted on the battlefield. This further implies that they must have the right to refuse an order which is determined to be unethical, and that they possess the ability to monitor and report on the ethical behavior of other military personnel as required.

### **Ethical Responsibilities**

I think of myself as a responsible scientist who has spent decades working on military applications of robotics. I think the following questions are crucial:

*Is it not our responsibility as scientists to look for effective ways to reduce human inhumanity to other people through technology? And if such inhumanity occurs during warfare, what can be done?*

It is my belief that research in ethical military robotics can and should be applied towards

achieving this end. But how can this happen? Where does humanity fit on the battlefield? Extrapolating these questions further, we ask:

#### **Should soldiers be robots?**

*Isn't that largely what they are trained to be?*

#### **Should robots be soldiers?**

*Could they be more humane than humans?*

One lesson I have learned along the way is that roboticists should not run from the difficult ethical issues surrounding the use of their intellectual property that is or will be applied to warfare, whether or not they directly participate. Wars unfortunately will continue and derivative technology from these ideas will be used. If your robotics research is of significance and it is published openly, it will be put to use in military systems by someone, somewhere, someday. Researchers are not immune from contributing to military applications by simply not accepting funds from the U.S. Department of Defense. To ensure proper usage of this technology, proactive management by all parties concerned is necessary. Complete relinquishment of robotics research as proposed by Bill Joy is the only alternative [12], but I do not personally favor that strategy.

I remain active in my research for the U.S. DOD in battlefield applications of robotics for both the U.S. Army and Navy regarding the deployment of teams of robots, but it remains a personal goal that these systems and other related military research products will ultimately be ethically restrained by technological methods such as those described in [13]-[15] so as to abide by the internationally agreed upon Laws of War. I also hope that this research will spur others into not only considering

this problem, but to help ensure that warfare is conducted justly, even with the advent of autonomous robots if international societies so deem it fit, and that those who step beyond those ethical bounds, whoever they may be, are successfully prosecuted for their war crimes. It is my conviction that as these weaponized autonomous systems appear on the battlefield, they should help to ensure that humanity, proportionality, responsibility, and relative safety, as encoded in the Laws of War, are extended during combat not only to friendly forces, but equally to noncombatants and those who are otherwise hors de combat, with the goal being a reduction in the loss of life of civilians and all other forms of collateral damage.

### **Author Information**

The author is with the Mobile Robot Laboratory, College of Computing, Georgia Institute of Technology, Atlanta, GA 30332-0280; arkin@cc.gatech.edu.

### **Acknowledgment**

This research is funded under Contract #W911NF-06-1-0252 from the U.S. Army Research Office. The author also would like to thank Patrick Ulam, Alan Wagner, and Brittany Duncan for their assistance in the research described in this paper.

### **References**

- [1] U.S. Surgeon General's Office, Mental Health Advisory Team (MHAT) IV Operation Iraqi Freedom 0507, Final Rep., Nov. 17, 2006.
- [2] H. Slim, *Killing Civilians: Method, Madness, and Morality in War*. New York, NY: Columbia Univ. Press, 2008.
- [3] D. Grossman, *On Killing: The Psychological Cost of Learning to Kill in War and Society*. Boston, MA: Little, Brown, 1995.
- [4] J. Bourke, *An Intimate History of Killing*. New York, NY: Basic, 1999.
- [5] R. Arkin, M. Fujita, T. Takagi, and R. Hasegawa, "An ethological and

- emotional basis for human-robot interaction," *Robotics and Autonomous Systems*, vol. 42, no. 3-4, Mar. 2003.
- [6] M. Walzer, *Just and Unjust Wars*, 4th ed. New York, NY: Basic, 1977.
- [7] S. Sagan, "Rules of engagement", in *Avoiding War: Problems of Crisis Management*, A. George, Ed.. Boulder, CO: Westview, 1991.
- [8] R. McLoughlin, "Fourth generation warfare and network-centric warfare," *Marine Corps Gazette*, 2006; <http://www.mca-marines.org/GAZETTE/06mcloughlin.asp>, accessed Sept. 11, 2008.
- [9] U.S. Defense Advanced Research Projects Agency (DARPA), *Scalable Network Monitoring*, Strategic Technology Office, Broad Agency announcement 07-52, Aug. 2007.
- [10] T. Adams, "Future warfare and the decline of human decision-making," *Parameters, U.S. Army War College Quart.*, pp. 57-71, Wint. 2001-2002.
- [11] M. Guetlein, "Lethal autonomous systems – Ethical and doctrinal implications," Naval War College Joint Military Operations Dept. pap., Feb. 2005.
- [12] B. Joy, "Why the future doesn't need us," *Wired*, no. 8.04, Nov. 2000.
- [13] R.C. Arkin, "Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture-Part I: Motivation and philosophy," in *Proc. Human-Robot Interaction 2008*, Amsterdam, NL, Mar. 2008.
- [14] R.C. Arkin, "Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture-Part II: Formalization for ethical control," in *Proc. 1st Conf. Artificial General Intelligence*, Memphis, TN, Mar. 2008.
- [15] R.C. Arkin, "Governing lethal behavior: Embedding ethics in a hybrid deliberative/reactive robot architecture-Part III: Representational and architectural considerations," in *Proc. Technology in Wartime Conf.*, Palo Alto, CA, Jan. 2008.