The Robot didn't do it.

A position paper for the
Workshop on Anticipatory Ethics, Responsibility and Artificial Agents

Ronald C. Arkin
School of Interactive Computing
Georgia Tech, Atlanta, GA 30332

This position paper addresses the issue of responsibility in the use of autonomous robotic systems. We are nowhere near autonomy in the philosophical sense, i.e., where there exists free agency and moral culpability for a non-human artificial agent. Sentient robots and the singularity are not concerns in the near to mid-term. While agents such as corporations can be held legally responsible for their actions, these exist of organizations under the direct control of humans. Intelligent robots, by virtue of their autonomous decision-making, are not of the same ilk.

For robots, as with any form of advanced technology, responsibility attribution must be made explicit and clear in their creation, manufacture, and deployment. Nowhere in the near future can I envision an intelligent artifact bearing responsibility for its own actions. Thus I advocate effective and informed acknowledgment of responsibility at all levels by humans prior to its use. For military systems deployed in the field, one possible method is through the use of a responsibility advisor, designed for both run-time and pre-mission aspects of a robotic mission. Details can be found in [1-3]. Regarding pre-deployment design, manufacturing, and policy decisions, responsibility attribution should be handled by proper certification and regulation prior to use.

I liken the problem of responsibility assignment for life-threatening errors committed by lethal autonomous robots during military operations to those of airborne precision-guided munitions landing on a school, hospital or religious structure. The question is: was this truly incidental in the performance of these duties and is the result covered by the Principle of Double Effect[1] should civilian casualties occur? Or was there negligence or deliberate malicious intent on the part of the designer, manufacturer, policymaker, commander, or warfighter? If so, then this atrocity must be documented effectively and the issue brought before the proper tribunals. Suitable system design can assist in evidence gathering for such prosecutions.

The question outstanding is whether or not existing International Humanitarian Law (IHL) adequately covers autonomous systems for military use. While there are those who argue supplemental laws or a treaty must be created to cover autonomous robots, the necessity is unclear. The debate should focus not on fear-mongering but rather where in current IHL is this technology inadequately covered. Specifically, if the hallmark criteria

---

1 The Principle (or Doctrine) of Double Effect, derived from the Middle Ages, asserts "that while the death or injury of innocents is always wrong, either may be excused if it was not the intended result of a given act of war" [5 (p. 258), 6]. As long as the collateral damage is an unintended effect (i.e., innocents are not deliberately targeted), it is excusable according to the Laws of War even if it is foreseen (and that proportionality is adhered to).

of humanity, military necessity, proportionality, and distinction derived from just war theory and enshrined in IHL are indeed lacking with respect to this new technology then suitable action must be taken. To date, there has been no convincing factual proof that that is the case.

Finally, the assertion that "warfare is an inherently human endeavor" will unfortunately not address the unrelenting plight of the non-combatant in the battlespace. If advanced technology can potentially focus on the persistent problem of the loss of innocent life during the commission of war crimes by humans, a humanitarian effort could result from the application of appropriate intelligent technology, possibly lowering (but not eliminating) atrocities in the battlefield. This is a worthwhile goal, and I would contend a responsibility that scientists who create and engineer advanced technology, including military robots, must address [4].

[1] Arkin, R.C., Wagner, A., and Duncan, B., "Responsibility and Lethality for Unmanned Systems: Ethical Pre-mission Responsibility Advisement", *Proc. 2009 IEEE Workshop on Roboethics*, Kobe JP, May 2009.

[2] Arkin, R.C. and Ulam, P., "Overriding Ethical Constraints in Lethal Autonomous Systems", Technical Report GIT-MRL-12-01, Georgia Tech Mobile Robot Laboratory, 2012.

[3] Arkin, R.C., Governing Lethal Behavior in Autonomous Robots, Chapman and Hall, 2009.

[4] Arkin, R.C., "Viewpoint: Military Robotics and the Robotics Community's Responsibility", *Industrial Robotics*, Vol. 38, No. 5, 2011.

[5] Wells, D., (Ed.), *An Encyclopedia of War and Ethic*s, Greenwood Press, 1996.

[6] Norman, R., *Ethics, Killing and War*, Cambridge University Press, Cambridge, England, 1995.