Lethality and Autonomous Robots: An Ethical Stance

Ronald C. Arkin
Georgia Institute of Technology

Battlefield ethics has for millennia been a serious question and constraint for the conduct of military operations by commanders, soldiers, and politicians, as evidenced for example by the creation of the Geneva conventions, the production of field manuals to guide appropriate activity for the warfighter in the battlefield, and the specific rules of engagement for a given military context.

Breeches in military ethical conduct often have extremely serious consequences, both politically and pragmatically, as evidenced recently by the Abu Ghraib and Haditha incidents in Iraq, which can actually be viewed as increasing the risk to U.S. troops there, as well as the concomitant damage to the United State's public image worldwide.

If the military keeps moving forward at its current rapid pace towards the deployment of intelligent autonomous robots, we must ensure that these systems be deployed ethically, in a manner consistent with standing protocols and other ethical constraints that draw from cultural relativism (our own society's or the world's ethical perspectives), deontology (right-based approaches), or within other related ethical frameworks.

Under the assumption that warfare unfortunately will continue into the foreseeable future in different guises, the question arises as to how will the advent of autonomous systems in the battlefield affect the conduct of war. There already exist numerous conventions, rules of war, military protocols, codes of conduct and rules of engagement, which are sometimes global in their application and at other times contextual, that are used to constrain of guide a human warfighter. Historically, mankind has been often unable to adhere to these rules/laws thus resulting in violations and war crimes.

Can autonomous systems do better? In this talk we study the underlying thesis that robots can ultimately be more humane than human beings in military situations, potentially resulting in a significant reduction of violations. This class of autonomous robots which maintain an ethical infrastructure to govern their behavior will be referred to as humane-oids.

A three year research effort on this topic is being conducted for the Army Research Organization, of which we are currently in the first year. Two topics are being investigated:

(1) What is acceptable? *Can we understand, define, and shape expectations regarding battlefield robotics.* A survey is being conducted to establish opinion on the use of lethality by autonomous systems spanning the public, researchers, policymakers, and military personnel to ascertain the current point-of-view maintained by various demographic groups on this subject.

(2) What can be done? *Artificial Conscience and Reflection*. We are designing a computational implementation of an ethical code within an existing autonomous robotic system, i.e., an "artificial conscience", that will be able to govern an autonomous system's behavior in a manner consistent with the rules of war.

The background and results to date of this research are presented in this talk.