

# Intelligent Agents in Military, Defense and Warfare: Ethical Issues and Concerns

Sahon Bhattacharyya  
Department of Computer Science  
St.Xavier's College (Autonomous),  
Kolkata, WB, India.  
Email: [sahon.dgro@acm.org](mailto:sahon.dgro@acm.org)

## ABSTRACT

Due to tremendous progress in digital electronics now intelligent and autonomous agents are gradually being adopted into the fields and domains of the military, defense and warfare. This paper tries to explore some of the inherent ethical issues, threats and some remedial issues about the impact of such systems on human civilization and existence in general. This paper discusses human ethics in contrast to machine ethics and the problems caused by non-sentient agents. A systematic study is made on paradoxes regarding the long-term advantages of such agents in military combat. This paper proposes an international standard which could be adopted by all nations to bypass the adverse effects and solve ethical issues of such intelligent agents.

KEYWORDS: *intelligent agents, military, computational ethics, problem of ethics, philosophy of science, artificial intelligence, machine ethics*

## I. INTRODUCTION

“I call upon the scientific community in the world, those who gave us nuclear weapons, to turn their great talents now to the cause of mankind and world peace: to give us the means of rendering these nuclear weapons impotent and obsolete.”  
—Ronald Reagan

As this decade-old millennium ushers in a new revolution of ever-advancing technology, attention inevitably turns to the novel idea of using this technology in the ancient and infamous fields of the military, defense and eventually, warfare. As with every new technology that has emerged from the intellect of man, we imagine the myriad ways in which we may apply it, from the benign domains of medical science to the useful needs of urban environments and remote explorations and eventually and inexorably to the more destructive needs of the battlefield. And it is this particular field which raises certain questions and issues, where we most need the humility of ethics.

*Ethics*, also known as *Moral Philosophy*, is a branch of philosophical inquiry that addresses questions about morality— that is, concepts such as good and evil, right and wrong, virtue and vice, justice, etc. *The Stanford Encyclopedia of Philosophy* states that we study ethics, as conceived by Aristotle, to improve our lives, and its principal concern is therefore the nature of human well-being.

Military, defense and warfare have always raised deep questions regarding right and wrong and have been the subject of innumerable debates worldwide. And as we hand over to humanity the gift of a modern technology, it is imperative that we pause to consider whether humanity benefits from it or turns it into an arsenal of destruction.

## II. WHY INTELLIGENT AGENTS?

Imagine the overwhelming sadness and irreparable loss of a soldier returning home in a flag-draped casket to a grief-stricken family (Lin, Bekey and Abney, 2008). It is quite understandable why the idea of machines on the battlefield appeals so much to the world. Machines do not die, they do not suffer pain. Not being sentient, they do not complain nor do they demand any rights like humans do. Being reconstructible, they offer a victory even in defeat.

The fear of death in human beings manifests itself as the desire to use machines endowed with intelligent technology in wars. The horror of death creates an aversion in the human mind to recoil at the idea of battle. Nonetheless it does in no way discourage the spirit of man from engaging in war. As a result, man readily sends machines to combat on his behalf, believing their temporal nature to be immaterial and in the pretence of saving countless human lives. Thus, such usage offers man a partial deliverance from the fear of death.

Intelligent machines, with high computational and decision making abilities also overcomes our limited capabilities, in effect, making them ‘smarter’ than human beings. Equipped with electronic memory and fast retrieval facility, it dexterously uses its knowledge database to figure out the optimal way to act, something which humans cannot do in a short period of time.

Intelligent machines are immune to and unaffected by emotions, adrenaline and stress, the factors which are detrimental to humans, causing them to overreact or overstep or even commit atrocities (Lin, Bekey and Abney, 2008). Not needing sleep, not requiring food, working continuously for very long periods of time, they offer very economically bright alternatives to sending humans to the battlefield.

### III. THE DOUBLE-EDGED SWORD!

Technology is a double-edged sword (Lin, Bekey and Abney, 2008). As much as the good it offers, inevitably with it comes evil. Intelligent agency is also not without its risks and threats. Some of these include (but are not limited to):

Unintentional or unlawful harm, brought upon by agents themselves unknowingly or based upon false or ambiguous information or knowledge about the environment or by the individuals controlling the agent, whether again, unknowably or with the intent to cause unlawful harm,

Possibility of serious malfunction at a critical time may occur and it since it has occurred in the past, it would be too naïve to consider that such a thing might not occur again; it thus becomes difficult to trust such a system with critical decision making especially when many lives are at stake,

Intelligent agents, at the end of the day, are governed by software systems not very unlike currently running our computers and thus, are very susceptible to capturing and hacking by unauthorized access. Malicious attacks such as these pose serious threats as the infiltrating party may decide to alter the programming in such a way that its actions are detrimental to its original cause, or may cause it to cause mass destruction, thereby turning it into a veritable weapon of terrorism, one of the very things we fight against!

We owe gratitude to science fiction for aiding in our imagination of a world with and without intelligent agencies in our day-to-day lives, the scenarios that would occur if intelligent and autonomous agents went out of control and wreaked havoc in our society, maybe even leading to human extermination.

### IV. LAWS OF ROBOTICS

Asimov (1950) discussed his famous three Laws of Robotics which he proposes should be followed by all intelligent and autonomous agents operating in our lives and environments:

*Law 1:* A robot may not injure a human being, or by inaction, cause a human being to come to any harm.

*Law 2:* A robot must obey orders given to it by human beings, except when such orders are in conflict with the First Law of Robotics.

*Law 3:* A robot must protect its own existence as long as it does not come into conflict with the First or the Second Law of Robotics.

A later addition was made (Asimov, 1985) called the Zeroth Law, which superseded all others:

*Law 0:* A robot may not harm humanity, or by inaction, allow humanity to come to any harm.

It may be asked that the First and the Zeroth Law are nearly identical but it is to be noted that there is a fundamental difference between them, namely, that if there ever was a choice to be made between protecting the whole of humanity and a single or a group of individuals, the preference would always go to the former. Thus, this addition entitles the agent to make sacrifices for the greater good.

But we still face many problems in this set of laws, one being the question of how an intelligent agent be made to recognize a situation of existential threat to humanity, thereby allowing it to harm individuals (Lin, Bekey and Abney, 2008).

An extended set of the Laws of Robotics was given by Roger Clarke (1994):

*Meta-Law:* A robot may not act unless its actions are subject to the Laws of Robotics.

*Law 0:* A robot may not harm humanity, or by inaction, allow humanity to come to any harm.

*Law 1:* A robot may not harm an individual, or by inaction, allow an individual to come to any harm, unless it violates a higher-order Law.

*Law 2:* A robot will always obey humans and its super-ordinate robots, unless in conflict with a higher-order Law.

*Law 3:* A robot must protect the existence of any super-ordinate robots and then itself, unless in violation of any higher-order Law.

*Law 4:* A robot must perform the duties it has been programmed to perform, unless such a task constitutes a violation of any higher-order Law.

*Procreation Law:* A robot may not take part in the design or manufacture of a robot unless the new robot's actions are subject to the Laws of Robotics.

But even in this system we see flaws. In the *Procreation Law*, what if the design of a new robot is aided by a law-abiding robot and later, the design of the new robot is altered without the assistance of the old robot and without letting it know? In this case, the new robot would have violated the Laws and with it, the actions of the old robot would also have been made illegal. This constitutes a paradox!

### V. MACHINE ETHICS

"*Machine ethics*, is concerned with adding an ethical dimension to machines. Unlike traditional computer ethics, which focuses on issues concerning human use of machines, machine ethics focuses on representing ethical principles explicitly and the challenge facing those working on machine ethics." (Anderson and Anderson, 2006)

This comes in two flavors: *implicit machine ethics* and *explicit machine ethics* (Moor, 2006). The former is where ethical principles have been formally programmed into the system of the intelligent agent, not quite unlike the Laws of Robotics discussed above. This provides a sort of safeguard

for intelligent agents which have been granted autonomy, at least in part.

The latter on the other hand presents more of a challenge. It would have to explicitly face ethical dilemmas and draw its own conclusion regarding ethical principles, represent them and act accordingly. The distinction, according to Anderson and Anderson (2007), lies not only in who is making the decision but also in the fact that such a system can justify its ethical principles.

The concern that these systems may start out with ethical behavior but end up behaving unethically is, according to Anderson and Anderson (2007), “stems from fears regarding human behavior” as humans are far from ideal systems. The authors even consider the possibility that such a system may grow to be even more ethical than humans themselves, an illustrious example being Andrew in Isaac Asimov’s science fiction novel, “The Bi-Centennial Man” (1976). (See Anderson and Anderson, 2007)

## VI. HUMAN ETHICS REVISITED

Even if in the distant unforeseeable future, if we, say, achieve the implementation of machine ethics – whether implicit or explicit, will it then be enough to allow the usage of such intelligent agents in the military?

As already mentioned above, humans are far from ideal and it is the nature of man to deviate from the standards of ethics towards pursuing his own interest, often at the cost of others. If machine ethics is achieved someday, it can be conjectured that the transient nature of human ethics will come into conflict with steadfast machine ethics and be detrimental to each other.

Humans are well-known for re-defining codes of ethics for their own purposes and come up with excuses for unethical behavior, which is something that cannot be expected of machines, as despite being intellectual fathomable, it is an idea well beyond current human ability.

Therefore, it becomes clear that the nature of human ethics and machine ethics are radically different and might not turn out to be completely compatible with each other and give rise to even more problems.

## VII. LACK OF HIGHER INTELLECT

In this section I would like to consider the so-called autonomy of ‘intelligent’ and ‘autonomous’ agents.

Autonomy is defined as the ability to make un-coerced and free decisions. In terms of movement and navigation and other decision making situations, such agents perform quite well, fulfilling their criterion for autonomy. But in the larger picture, where we have envisioned ethical agents, surely we can expect such agents to be against wars and

military conflicts on grounds that they are harmful to humanity in the long run.

But again, if they are programmed so, to act according to the instructions given to them, I find it hard to believe that they have any autonomy at all. At most they can be said as having partial autonomy in their decisions, lacking intentionality and free will, which in my opinion is crucial for autonomy.

This can be attributed to the lack of intellect on part of the intelligent agents. They are merely the carriers of the autonomy of their makers and conceivers – bearing the will and intention of war. And if machine ethics were to be incorporated we might find them refusing them to participate in any military conflict at all. And this might logically lead to military organizations around the world removing the safeguarding device that implements machine ethics, which brings us promptly back to square one, i.e., where we started from – at the conflict of human and machine ethics.

Thus, we find that the issues of ethical behavior in intelligent agents lead to deceptive paradoxes – which we must rid ourselves of before committing ourselves to using such agents in the military, defense and warfare. As a result, we might discover that the problem of ethics is unsolvable, intractable or that it might not yield an optimal solution at all and their usage might prove to be detrimental in everyway.

## VIII. THE HUMANE SIDE OF DEATH

As we discussed in Section II, one of the main reasons behind the introduction of intelligent machinery in military was to prevent human death and casualty. But now I propose to consider the question whether, in the very long run, this ‘intelligent’ solution really achieves a better end than death does?

Let us recall the World Wars I and II. After the wars ended, the horrifying and appalling damage caused by the wars caused all nations to join efforts in the making of the United Nations to prevent such wars in future and to promote international cooperation, social progress, human rights, and harmony and world peace.

Mortality and the fear of death play a vastly significant role in the shaping of man’s mind and life. Humans are the way they are because of the fear of death. It is this fear which makes us humane and thus, human. It is this fear which brings into humans their compassion, mercy and dominating will for world peace. It is this *sapience* which make distinct the *homo sapiens*, something machine agents can never achieve, however intelligent and autonomous they may be.

Man needs death to remind him of his mortality and to make him work towards attaining world peace. If intelligent agents were allowed to replace humans in wars in an attempt to ‘save’ human lives we might find that paradigms of world peace might start changing in some undesirable way.

Compassion would no longer play a role and the motivation for world peace will be subject to significant decline.

Therefore, the question is which one do we choose? To let humans fight in wars, to let them die so that others be compassionate and work towards slowly but surely towards eventual world peace? Or to let machines take over and compel us to abandon world peace policies and take joy in the fact that human lives are being saved as the world incessantly battles on?

#### IX. WHAT CAN BE DONE?

We cannot deny that the military intelligence organizations require technology to move forward and yet we are faced with the dilemmas of harmful and adverse effects on humanity. Then how do we reconcile these two needs without abandoning one in favor of the other?

As a tentative solution, I propose that international standards be adopted and agreed upon by all nations that the use of intelligent and autonomous agents be restricted to intra-national affairs only and that violations of this standard be made illegal and punishable.

All nations may not be able to develop intelligent machines of the same capabilities. And this might lead inexorably to a veritable *arms race* among nations, sowing dissension and the seeds of conflict, leading to war.

#### X. CONCLUSION

“In the end, war, no matter how necessary, or how justified, remains a crime!” —Ernest Hemingway

It is to be remembered that scientific inquiry, investigation and exploration exists for the better future of humanity and the sustenance of human civilization.

As the message of Ronald Reagan, at the beginning of this article, calls out to the scientific community to render impotent and obsolete the weapons of destruction, I would like to point out that we need not turn great talents to achieving the above humane aim because ethics and ethics alone serves the noble purpose envisioned therein. It is in ethics therefore that we must place our trust and our hopes of a better world.

And therefore, before we unveil every new technology, however altruistic in nature, and present it to the world, it becomes our solemn duty as scientists, humble seekers after truth and guardians of a better tomorrow, to ask ourselves of the knowledge that we hold in our hands: Shall it redeem us, or shall it condemn us?

#### REFERENCES

- [1] Patrick Lin, George Bekey, Keith Abney, 2008: “Autonomous Military Robotics: Risk, Ethics and Design” (US Department of Navy, Office of Naval Research, under awards #N00014-070101152 and #N00014-08-1-1209)
- [2] Isaac Asimov, 1950: “I, Robot”
- [3] Tuncer Oren: “Ethics as a basis for sustainable civilized behaviour for human and software agents”, Acta Systemica, 2:1, 1-5
- [4] Michael Anderson and Susan Leigh Anderson: “Machine Ethics: Creating an Ethical Intelligent Agent”, AI Magazine, Vol.28 No. 4 (2007)