

On the Ethical Quandaries of a Practicing Robotist: A first-hand look

Ronald C. Arkin
Mobile Robot Laboratory
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332, U.S.A.

Abstract. Robotics has progressed substantially over the last 20 years, moving from simple proof-of-concept experimental research to developing market and military technologies that have significant ethical consequences. This paper provides the reflections of a robotist on current research directions within the field and the social implications associated with its conduct.

Keywords: robot ethics, military robotics, entertainment robotics

Introduction

I have been a practitioner in the field of robotics for over 20 years, and during that time I developed a strong appreciation for the potential ramifications of the research that I have been and currently am conducting, ranging from the purely scientific to the more applied. This has led me to delve deeply into the questions surrounding the ethical practice of robotics as a whole and to seek out the means for analysis of the consequences of my personal actions in the field (past, present, and future) while also actively encouraging my colleagues to do so.

There are all sorts of red flags being raised by others regarding the perils of robotics, all the way from a predicted end of the human-dominated world due to self-replicating robots (e.g., [1,2]) to far more immediate issues surrounding the application of robotics (e.g., the use of robots in warfare [3-5], labor ramifications, and the deliberate psychological manipulation of human beings by robot entities [6-8]). While I could also take a stand on the more alarmist perspectives, I will in this article, address those concerns arising from the here-and-now practice of robotics from a personal perspective, most of which have serious short-term ethical consequences. While some of these issues have been discussed in prior Roboethics conferences (e.g., [9,10]) in a more general informative manner, they have not been developed in the context of an individual researcher's perspective nor, oddly enough, in a true ethical context, where different theories of ethical reasoning are applied, whether they be utilitarian, cultural relativism, social contract theory, Kantian, etc.

Independent of the specific personal ethical framework for analysis chosen, I will lay in front of you three ethical quandaries that are not hypothetical but constitute the reality that I have been or am currently confronted with. In teaching my class on robot

ethics (CS 4002 Robots and Society) I encourage my students to examine not only abstract or removed case studies but also current practices such as my own in light of the criticisms they may well be subjected to within society. I find this exercise invaluable personally as well, as it informs me, often in surprising ways regarding the views that at least one segment of the population holds regarding this research.

Quandary 1: Autonomous Robots Capable of Lethal Action

One major research area I am responsible for involves military robotics. While I choose to only conduct unclassified research so that I can publish and talk freely about my work (at least to date), my experience in this area ranges from areas such as robots for explosive ordnance disposal and humanitarian demining to the development of software for autonomous weapons-bearing unmanned vehicle systems (e.g., the Defense Advanced Research Project Agency's (DARPA) Unmanned Ground Combat Program). The controversy surrounding this application is clearly evident, ranging from the traditional arguments against warfare in general and new weapon construction in particular, to issues surrounding the direct application of lethality by autonomous systems without having a human in direct control or issuing a confirmation of an order to kill. Ongoing research on my part for the U.S. Army involves assaying opinion (of the public, researchers, the military, and policymakers) on the use of this latter class of autonomous robots, while also investigating how to embed an "artificial conscience" in these vehicles to ensure that the international laws of war and rules of engagement are strictly followed by machines, perhaps even more effectively than by humans. This has required developing an understanding of Just War theory [11] and delineating methods by which combatant/noncombatant discrimination, proportionality of force, minimization of collateral damage and unnecessary suffering, and related *Jus in Bello* ethical issues can be enforced within autonomous robots. This research speaks predominantly to the deontological basis as encoded in International Conventions (e.g., Hague and Geneva Protocols) in addition to utilitarian considerations in terms of military necessity, weapon selection, firing pattern, and permission to fire, all subject to the former rights-based restrictions.

As it is clear to me that the technology that I helped create, specifically autonomous robotic architectures, is moving forward in warfare applications, with or without my participation, I feel compelled to act in a manner that leads to the development of autonomous systems that are capable of complying with accepted International Law. Further I have become convinced of the highly controversial position, due to the propensity of soldiers to tolerate and commit illegal acts under wartime conditions [12], that ultimately autonomous systems can outperform humans ethically in the battlefield. Details of this ethics-driven approach can be found in [5].

Quandary 2: Entertainment Robotics and the Suspension of Reality

The second area of ethical controversy deals with personal robotics. I have served as a consultant for Sony Corporation for nearly 10 years in the development of software for the AIBO and QRIO entertainment robots (Figure 1) [13]. While most researchers view this activity as an innocuous and even beneficial use of robotics, possibly for the treatment of isolated elderly people, not all agree [6,7]. This research requires a deep

understanding of not only a robot's capabilities but also human psychology, where the roboticist's deliberate goal is to induce pleasant psychological states in the observer through specialized patterns of robot behavior and, to the greatest extent possible, suspend observer disbelief that this robot is not alive. The intended goal is to establish a long-term, even lifelong, human-robot relationship, perhaps even a dependency not unlike what is experienced with pets or among friends.

Some view the ethics for this type of research as no different than that of advertising, cinema, video games, or other forms of entertainment. Others such as Sparrow [6] argue that this is an intrusion into the rights of the elderly to remain in contact with the real world, while society (and researchers such as myself) makes excuses for its intended unethical use. Robotics researchers make a tacit assumption that the creation of this new technology is wholly appropriate and can only enrich the lives of those on the receiving end. It is important that we as scientists re-examine this assumption. To that end, I have gone to the heart of the scientific source [14] to present contradictory perspectives based on deontological arguments and potential violations of the social contract that challenge the underlying assumption that the goal of this form of robotics research is ethically acceptable.



Figure 1: QRIO is the humanoid on the left; AIBO is the robot dog on the right.

Quandary 3: Robotics and Unemployment

The final area of personal concern involves the displacement of workers, in areas such as shipyards. Although I currently have limited current research in this area, I am considering and willing to expand it, which in many respects has caused me more soul-searching than the other two examples cited above. I have also conducted research

extensively in the applications of autonomous robotics to manufacturing in years past [15,16].

Indeed much of the underlying premise for the use of robotics as a whole is the elimination of the three D jobs: those that are Dull, Dangerous, and Dirty. While this at first blush appears to be a noble goal, without concomitant social support we are just encouraging the same forms of social upheaval that accompanied the earlier industrial revolution. From a corporate perspective, this research avenue can undoubtedly lead to a clash between an act utilitarian perspective of a large industrial concern with the individual worker's (Kantian) right to good will.

When a roboticist can project the consequences of their research as ultimately leading to significant unemployment with worldwide impact, and while being unable to directly influence social support structures for those potentially made unemployed, what is their moral responsibility here? This may lead to a more traditional debate on industrial revolutions in general, but nonetheless roboticists often are woefully unaware of where the consequences of their work may lead in this domain.

In Summary

These issues are personal day-to-day concerns, and I contend they should also be part of a regular professional roboticist's diet. As in many ethical areas, we will not agree universally on the outcomes for these and other related issues, at least from an individual perspective. Nonetheless, I argue that it is a central responsibility of a roboticist to conduct such self-examinations to ensure that he/she is aware, at least consistent with their own morality, the consequences of their actions and also to be prepared to become engaged with others in this field on related ethical concerns, so that we as a group of concerned scientists can develop acceptable limits and guidelines to a broad range of emerging robotics issues. Reaching out and engaging others from non-technical communities such as philosophers, social and political scientists is crucial toward achieving this end.

Acknowledgment

A portion of this research is funded under Contract #W911NF-06-0252 from the U.S. Army Research Office.

References

- [1] Joy, William, "Why the Future Doesn't Need Us", *Wired*, Issue 8.04, April 2000.
- [2] Moravec, Hans, *Mind Children: The Future of Robot and Human Intelligence*, Harvard University Press, 1990.
- [3] Sparrow, R., "Killer Robots", *Journal of Applied Philosophy*, Vol. 24, No.1, 2006.
- [4] Asaro, P., "How Just Could a Robot War Be?" presentation at *5th European Computing and Philosophy Conference*, Twente, NL June 2007.

- [5] Arkin, R.C., "Governing Lethal Behavior: Embedding Ethics in a Hybrid Deliberative/Reactive Robot Architecture", Technical Report GIT-GVU-07-11, College of Computing, Georgia Institute of Technology, 2007.
- [6] Sparrow, Robert, "The March of the Robot Dogs", *Ethics and Information Technology*, Vol. 4 No. 4, 2002, pp. 305-318.
- [7] Sparrow, R. and Sparrow, L., "In the Hands of Machines? The Future of Aged Care", *Mind and Machines*, Vol. 16, pp. 141-161, 2006.
- [8] Krahling, M., "In Between Companion and Cyborg: The Double Diffracted Being Elsewhere of a Robodog", *International Review of Information Ethics*, Vol. 6, pp. 69-77, December 2006.
- [9] First International Symposium on Roboethics, Villa Nobel, San Remo Italy, January 2004. http://www.roboethics.org/sanremo04/ROBOETHICS_Program.html.
- [10] Proceedings of Roboethics Workshop at 2007 IEEE International Conference on Robotics and Automation, Rome, Italy, April 2007. <http://www.roboethics.org/icra07/contributions.html>.
- [11] Walzer, M., *Just and Unjust Wars*, 4th Ed., Basic Books, 1977.
- [12] Surgeon General's Office, Mental Health Advisory Team (MHAT) IV Operation Iraqi Freedom 05-07, Final Report, Nov. 17, 2006.
- [13] Arkin, R., Fujita, M., Takagi, T., and Hasegawa, R., "An Ethological and Emotional Basis for Human-Robot Interaction", *Robotics and Autonomous Systems*, 42 (3-4), March 2003.
- [14] Arkin, R.C., "Ethical Issues Surrounding the Use of Robotic Companions for the Elderly: Illusion versus Reality", presentation at *Workshop on Assistive Technologies: Rehabilitation and Assistive Robotics*, held at *IEEE/RSJ 2007 International Conference on Intelligent Robotics and Systems (IROS '07)*, San Diego, CA, October, 2007.
- [15] Arkin, R.C. and Murphy, R.R., "Autonomous Navigation in a Manufacturing Environment", *IEEE Transactions on Robotics and Automation*, Vol. 6, No. 4, pp. 445-454, August 1990.
- [16] Murphy, R. and Arkin, R.C., "Autonomous Mobile Robots in Flexible Manufacturing Systems", *Proc. Fourth International Conference on Artificial Intelligence Applications*, San Diego, CA, pp. 412-414, 1988.