

# Sharing Moral Responsibility with Robots: A Pragmatic Approach

Gordana DODIG-CRNKOVIĆ<sup>1</sup> and Daniel PERSSON<sup>2</sup>

*School of Innovation, Design and Engineering,  
Mälardalen University, Västerås, Sweden*

**Abstract.** Roboethics is a recently developed field of applied ethics which deals with the ethical aspects of technologies such as robots, ambient intelligence, direct neural interfaces and invasive nano-devices and intelligent soft bots. In this article we look specifically at the issue of (moral) responsibility in artificial intelligent systems. We argue for a pragmatic approach, where responsibility is seen as a social regulatory mechanism. We claim that having a system which takes care of certain tasks intelligently, learning from experience and making autonomous decisions gives us reasons to talk about a system (an artifact) as being “responsible” for a task. No doubt, technology is morally significant for humans, so the “responsibility for a task” with moral consequences could be seen as moral responsibility. Intelligent systems can be seen as parts of socio-technological systems with distributed responsibilities, where responsible (moral) agency is a matter of degree. Knowing that all possible abnormal conditions of a system operation can never be predicted, and no system can ever be tested for all possible situations of its use, the responsibility of a producer is to assure proper functioning of a system under reasonably foreseeable circumstances. Additional safety measures must however be in place in order to mitigate the consequences of an accident. The socio-technological system aimed at assuring a beneficial deployment of intelligent systems has several functional responsibility feedback loops which must function properly: the awareness and procedures for handling of risks and responsibilities on the side of designers, producers, implementers and maintenance personnel as well as the understanding of society at large of the values and dangers of intelligent technology. The basic precondition for developing of this socio-technological control system is education of engineers in ethics and keeping alive the democratic debate on the preferences about future society.

**Keywords.** Intelligent agents, Moral responsibility, Safety critical systems

## Introduction

Engineering can be seen as a long-term, large-scale social experiment since the design, production and employment of engineered artifacts can be expected to have long-range effects [1]. Especially interesting consequences might be anticipated if the engineered artifacts are intelligent, adaptive and autonomous. Recently, Roboethics, a field of

---

<sup>1</sup> gordana.dodig-crnkovic@mdh.se

<sup>2</sup> dpn04001@student.mdh.se

applied ethics, has developed with many interesting, novel insights.<sup>3</sup> Topics addressed within Roboethics include the use of robots, ubiquitous sensing systems and ambient intelligence, direct neural interfaces and invasive nano-devices, intelligent soft bots, robots aimed at warfare, and similar, which actualize ethical issues of responsibility, liability, accountability, control, privacy, self, (human) rights, and similar [2].

This article deals specifically with the issue of (moral) responsibility in artificial intelligent systems. We argue that this should be handled by adopting a pragmatic approach, where responsibility is seen as a social regulatory mechanism.

## 1. Moral Responsibility and Intelligent Systems

Moral responsibility is understood as consisting of two parts: causal responsibility and intention. Traditionally only humans are considered to be capable of the mental state of intention. This mental state can be seen as the origin of an act that, depending on the effects it causes, can imply moral responsibility [3][4].

A common argument against ascribing moral responsibility to artificial intelligent systems is that they are not considered to have the capacity for mental states like intention [3][4]. Another argument maintains that it is pointless to assign praise or blame to an agent of this type when it has no meaning to the agent [5].

Both these arguments stem from a view in which agents are seen primarily as isolated entities. Dennett and Strawson suggest that we should understand moral responsibility not as individual duty, but as *a role defined by externalist pragmatic norms of a group* [6][7]. In this functionalist view moral responsibility can best be seen as a social regulatory mechanism which aims at enhancing actions considered to be good, and simultaneously minimizing what is considered to be bad.

We argue that to address the question of ascribing moral responsibility to intelligent systems we must adopt the functionalist view and see them as parts of larger socio-technological systems with distributed responsibilities, where responsibility of a moral agent is a matter of degree. From such a standpoint ascribing responsibility to an intelligent system has primarily a regulatory role. Delegating a task to a machine is also delegating responsibility for the safe and successful completion of that task to the machine [8]. A machine that takes care of certain tasks intelligently, learning from experience and making autonomous decisions gives us good reasons to talk about a machine as being “responsible” for a task in the same manner that we talk about a machine being “intelligent”. No doubt, technology is morally significant for humans, so the “responsibility” for a task with moral consequences could be seen as moral responsibility. Moral responsibility as a regulative mechanism shall not only locate the blame but more importantly assure future appropriate behavior of the system. *Consequential responsibility*, which presupposes *moral autonomy*, will however be distributed through the system.

---

<sup>3</sup> See, e.g. <http://www.roboethics.org>, <http://roboethics.stanford.edu/>, or *International Review of Information Ethics* Vol. 6, IRIE, 2006, that was dedicated to Ethics of Robotics, <http://www.i-r-i-e.net/archive.htm>

## 2. Risks and Distribution of Responsibility in Intelligent Technology

Based on the experiences with safety critical systems such as nuclear power, aerospace and transportation systems one can say that the socio-technological structure which supports their beneficial functioning is a system of safety barriers preventing and mitigating malfunction. The most important part is to assure safe functioning under normal conditions, which is complemented by the supposed abnormal/accidental condition scenarios. There must be several levels of organizational and physical barriers in order to cope with different levels of severity of malfunctions [9].

In every design process there are uncertainties that are the result of our limited resources. All new products are tested under certain conditions in a given context. This implies that an engineered product may, sooner or later, in its application be used under conditions for which it has never been tested. Even in such situations we expect the product to function safely. Handling risk and uncertainty in the production of a safety critical technical system is done on several levels. Producers must take into account everything from technical issues, through issues of management and organization, to larger issues on the level of societal impact [10]. Risk assessment is a standard way of dealing with risks in the design and production of safety critical systems [11], also relevant for intelligent systems.

Any technology subject to uncertainty and with a potentially high impact on human society is expected to be handled cautiously, and intelligent systems surely fall into this category. Thus, preventing harm and having the burden of proof of harmlessness is something that producers of intelligent systems are responsible for. (Precautionary Principle<sup>4</sup>)

A precondition for this socio-technological control system is an engineer informed about the ethical aspects of engineering, where education in professional ethics for engineers is a fundamental factor [12].

## 3. Conclusion

According to the classical approach, free will is essential for an agent to be assigned moral responsibility. Pragmatic approaches on the other hand focus on social, organizational and role-assignment aspects of responsibility. We argue that moral responsibility in intelligent systems is best viewed as a regulatory mechanism, and follow essentially a pragmatic (instrumental, functionalist) line of thought. Intelligent systems can be seen as parts of socio-technological systems with distributed responsibilities, where responsible (moral) agency is a matter of degree. We claim that for all practical purposes, the question of responsibility in safety critical intelligent systems may be addressed in the same way as the safety in traditional safety critical systems, such as nuclear industry and transports.

Long-term, wide range consequences of the deployment of intelligent systems in human societies must be discussed on a democratic basis as the intelligent systems have a potential of radically transforming the future of humanity. Education in professional ethics for engineers is a fundamental factor for building a socio-technological system of responsibility.

---

<sup>4</sup> [http://ec.europa.eu/dgs/health\\_consumer/library/pub/pub07\\_en.pdf](http://ec.europa.eu/dgs/health_consumer/library/pub/pub07_en.pdf)

## References

- [1] Martin, M.W., Schinzinger, R., *Ethics in Engineering*, McGraw-Hill, 1996.
- [2] Dodig-Crmkovic G., Professional Ethics in Computing and Intelligent Systems, *Proceedings of the Ninth Scandinavian Conference on Artificial Intelligence (SCAI 2006)*, Espoo, Finland, October 25-27, 2006.
- [3] Johnson D. G., Computer systems: Moral entities but not moral agents, *Ethics and Information Technology*, Vol. 8, Springer, 2006, pp. 195-204.
- [4] Johnson D. G. and Miller K. W., A dialogue on responsibility, moral agency, and IT systems, *Proceedings of the 2006 ACM symposium on Applied computing table of content*, Dijon, France, 2006, pp. 272 – 276.
- [5] Floridi L. and Sanders J. W., On the morality of artificial agents, *Minds and Machines*, Vol. 14, Kluwer Academic Publishers, 2004, pp. 349-379.
- [6] Dennett, D. C., Mechanism and Responsibility, in *Essays on Freedom of Action*, T. Honderich (ed), Routledge & Keegan Paul, Boston, 1973.
- [7] Strawson P. F., Freedom and Resentment, in *Freedom and Resentment and Other Essays*, Methuen, 1974.
- [8] Johnson D. G. and Powers T. M., Computer systems and responsibility: A normative look at technological complexity, *Ethics and Information Technology*, Vol. 7, Springer, 2005, pp. 99-107.
- [9] Dodig-Crmkovic G., ABB Atom's Criticality Safety Handbook, ICNC'99 Sixth International Conference on Nuclear Criticality Safety, Versailles, France, (1999), Available: <http://www.idt.mdh.se/personal/gdc/work/csh.pdf>
- [10] Huff, C., Unintentional Power in the Design of Computing Systems, in T. W. Bynum and S. Rogerson, eds., *Computer Ethics and Professional Responsibility*, Blackwell Publishing, Kundli, India, 2004, pp. 98-106.
- [11] Stamatelatos M., Probabilistic Risk Assessment: What Is It And Why Is It Worth Performing It?, NASA Office of Safety and Mission Assurance, 2000, Available: <http://www.hq.nasa.gov/office/codeq/qnews/prs.pdf>
- [12] Dodig-Crmkovic G., On the Importance of Teaching Professional Ethics to Computer Science Students, Computing and Philosophy Conference, E-CAP 2004, Pavia, Italy, in L. Magnani, ed., *Computing and Philosophy*, Associated International Academic Publishers, 2005.