Towards Trustworthy Intelligent Robots -
A Pragmatic Approach to Moral Responsibility

Gordana Dodig-Crnkovic & Daniel Persson
School of Innovation, Design and Engineering, Mälardalen University Sweden

http://www.idt.mdh.se/personal/gdc
Gordana Dodig-Crnkovic and Daniel Persson
Tenth Scandinavian Conference on Artificial Intelligence SCAI 2008.
Volume 173, Frontiers in Artificial Intelligence and Applications.
Eds. A. Holst, P. Kreuger and P. Funk, 2008
1. Introduction
As artificially intelligent artifacts and systems become increasingly independent, adaptive, and autonomous it is important to reconsider issues of responsibility for decisions made by such artifacts and systems.
We propose an approach where responsibilities are divided among agents in a complex socio-technological system, where responsible (moral) agency is a matter of degree.

2. Roboethics
Roboethics is a newly established field of applied ethics that deals with the ethical implications of production, use, and interaction with artificially intelligent artifacts and systems such as:
- Robots
- Softbots
- Direct neural interfaces
- Ubiquitous sensing systems

3. Views of responsibility
In the traditional view, normal responsibility consists of intention and causality. An agent that caused an event and intended for it to happen is considered responsible for the event.
This view of responsibility is focused on the individual agent. Applying this view to responsibility for artificial agents is controversial since these do not recognize the capability of internal states of intention.

4. Distribution of responsibility
The production and use of modern technological artifacts and systems in most cases require a complex distribution and delegation of responsibilities.

5. Risks and responsibility
In every design process there are uncertainties caused by limited awareness, products may not be used under conditions they have not been designed or tested for.
Safety critical intelligent systems must therefore be supported by socio-technological structures with different levels of organisational and physical barriers set up to secure their safe function under different conditions.

6. Conclusion
Intelligent systems are parts of socio-technological systems, where responsible (moral) agency is distributed, and a matter of degree based on functional relationships.
Practically, questions of responsibility in such systems may be addressed in the same way as in classical socio-technological systems.
A fundamental factor for establishing effective socio-technological systems of responsibility is education in professional ethics for engineers, and a broad, democratically based discussion on the production and use of intelligent systems.
Roboethics

www.roboethics.org
Information Ethics
Moral agency
Views of responsibility
Classical view

Causality
+
intention
Pragmatic view

Regulatory mechanism
Socio-technological network

Delegation & distribution
Responsible agents
“natural” & “artificial”
Safety barriers

- Physical
- Organizational
Artifactual intelligence: behavior that in human would require intelligence.
Artifactual morality: behavior that in human would require morality.
RESPONSIBILITY
ASCRIPITION

Artifacts ascribed artifactual responsibility for a task in planning operations
Proactive approach

- prediction,
- prevention,
- preparedness.
Legal aspects

analysis of technosocial system done on case by case basis.
CONVERGENCE