



# Machine Ethics

Papers from the AAI Fall Symposium

Technical Report FS-05-06



AAAI Press

American Association for Artificial Intelligence

**AAAI Press**

445 Burgess Drive  
Menlo Park, California 94025

ISBN 978-1-57735-252-5 FS-05-06



# Machine Ethics

Papers from the AAI Fall Symposium

*Michael Anderson, Susan Leigh Anderson, and Chris Armen, Cochairs*

Technical Report FS-05-06

AAAI Press  
Menlo Park, California

Copyright © 2005, AAAI Press

The American Association for Artificial Intelligence  
445 Burgess Drive  
Menlo Park, California 94025 USA

AAAI maintains compilation copyright for this technical report and retains the right of first refusal to any publication (including electronic distribution) arising from this AAAI event. Please do not make any inquiries or arrangements for hardcopy or electronic publication of all or part of the papers contained in these working notes without first exploring the options available through AAAI Press and *AI Magazine* (concurrent submission to AAAI and another publisher is not acceptable). A signed release of this right by AAAI is required before publication by a third party.

ISBN 978-1-57735-252-5

FS-05-06

Manufactured in the United States of America

# Organizing Committee

Michael Anderson (Cochair), University of Hartford (anderson@hartford.edu)  
Susan Leigh Anderson (Cochair), University of Connecticut (susan.anderson@uconn.edu)  
Chris Armen (Cochair), Trinity College (chris.armen@trincoll.edu)

This AAAI Symposium was held November 4–6, 2005,  
in Crystal City, Arlington, Virginia USA



# Contents

Preface / vii

*Michael Anderson, Susan Leigh Anderson, and Chris Armen*

## Papers

Towards Machine Ethics: Implementing Two Action-Based Ethical Theories / 1

*Michael Anderson, Susan Leigh Anderson, and Chris Armen*

Asimov's "Three Laws of Robotics" and Machine Metaethics / 8

*Susan Leigh Anderson*

Toward Ethical Robots via Mechanized Deontic Logic / 17

*Konstantine Arkoudas, Selmer Bringsjord, and Paul Bello*

Ethical Robots: The Future Can Heed Us / 24

*Selmer Bringsjord*

Additional Thoughts Concerning the Legal Status of a Non-Biological Machine / 30

*David J. Calverley*

The Utilibot Project: An Autonomous Mobile Robot Based on Utilitarianism / 38

*Christopher Cloos*

There is no 'I' in 'Robot': Robotic Utilitarians and Utilitarian Robots / 46

*Christopher Grau*

Particularism and Generalism: How AI Can Help Us to Better  
Understand Moral Cognition / 52

*Marcello Guarini*

Technological Artifacts as Moral Carriers and Mediators / 62

*Lorenzo Magnani*

Lessons in Machine Ethics from the Perspective of Two  
Computational Models of Ethical Reasoning / 70

*Bruce M. McLaren*

The Nature and Importance of Machine Ethics / 78

*James H. Moor*

Deontological Machine Ethics / 79

*Thomas M. Powers*

What Statistics Could Do for Ethics?  
The Idea of Common Sense Processing Based Safety Valve / 85  
*Rafal Rzepka and Kenji Araki*

A Robust View of Machine Ethics / 88  
*Steve Torrance*

Machine Morality: Bottom-up and Top-down Approaches  
for Modeling Human Moral Faculties / 94  
*Wendell Wallach, Colin Allen, and Iva Smit*

### **Poster Abstracts**

Machine Ethics and Human Ethics: A Critical View / 103  
*Francesco Amigoni and Viola Schiaffonati*

The Paradox of Social Robotics: A Discussion / 105  
*Brian R. Duffy and Gina Jone*

Moral Intelligence for Human and Artificial Agents / 107  
*Laura Pana*

The Ambiguous Ethical Status of Autonomous Robots / 109  
*John P. Sullins*



# Preface

Past research concerning the relationship between technology and ethics has largely focused on responsible and irresponsible use of technology by human beings, with a few people being interested in how human beings ought to treat machines. In all cases, only human beings have engaged in ethical reasoning. The time has come for adding an ethical dimension to at least some machines. Recognition of the ethical ramifications of behavior involving machines, as well as recent and potential developments in machine autonomy, necessitate this. In contrast to computer hacking, software property issues, privacy issues and other topics normally ascribed to computer ethics, machine ethics is concerned with the behavior of machines towards human users and other machines.

We contend that research in machine ethics is key to alleviating concerns with autonomous systems—it could be argued that the notion of autonomous machines without such a dimension is at the root of all fear concerning machine intelligence. Further, investigation of machine ethics could enable the discovery of problems with current ethical theories, advancing our thinking about Ethics. This symposium brought together participants from a wide variety of disciplines to the end of forging a set of common goals for machine ethics investigation and the research agendas required to accomplish them. Topics included:

- improvement of interaction between artificially and naturally intelligent systems through the addition of an ethical dimension to artificially intelligent systems
- enhancement of machine-machine communication and cooperation through an ethical dimension
- design of systems that provide expert guidance in ethical matters
- deeper understanding of ethical theories through computational simulation
- development of decision procedures for ethical theories that have multiple prima facie duties
- computability of ethics
- theoretical and practical objections to machine ethics
- impact of machine ethics on society

The symposium was comprised of invited speakers, paper and poster presentations, and system demonstrations. It was an interesting, informative, and productive exploration of this important emerging topic.

Further information regarding machine ethics can be found at <http://www.MachineEthics.org>.

*Michael Anderson, Susan Leign Anderson, and Chris Armen*