MACHINE ETHICS

The new field of machine ethics is concerned with giving machines ethical principles, or a procedure for discovering a way to resolve the ethical dilemmas they might encounter, enabling them to function in an ethically responsible manner through their own ethical decision making. Developing ethics for machines, which can be contrasted with developing ethics for human beings who use machines, is by its nature an interdisciplinary endeavor.

The essays in this volume represent the first steps by philosophers and artificial intelligence researchers toward explaining why it is necessary to add an ethical dimension to machines that function autonomously, what is required in order to add this dimension, philosophical and practical challenges to the machine ethics project, various approaches that could be considered in attempting to add an ethical dimension to machines, work that has been done to date in implementing these approaches, and visions of the future of machine ethics research.

Dr. Michael Anderson is a Professor of Computer Science at the University of Hartford, West Hartford, Connecticut. His interest in further enabling machine autonomy led him first to investigate how a computer might deal with diagrammatic information – work that was funded by the National Science Foundation – and has currently resulted in his establishing machine ethics as a bona fide field of scientific inquiry with Susan Leigh Anderson. He maintains the Machine Ethics Web site (http://www.machinethics.org).

Dr. Susan Leigh Anderson is Professor Emerita of Philosophy at the University of Connecticut. Her specialty is applied ethics, most recently focusing on biomedical ethics and machine ethics. She has received funding from the National Endowment for the Humanities and, with Michael Anderson, from NASA and the NSF. She is the author of three books in the Wadsworth Philosophers Series, as well as numerous articles.

Together, the Andersons co-chaired the AAAI Fall 2005 Symposium on Machine Ethics, co-edited an IEEE Intelligent Systems special issue on machine ethics, and co-authored an invited article on the topic for Artificial Intelligence Magazine. Their research in machine ethics was selected for Innovative Applications of Artificial Intelligence as an emerging application in 2006, and the October 2010 issue of Scientific American Magazine featured an invited article on their research in which the first robot whose behavior is guided by an ethical principle was debuted.
Machine Ethics

Edited by
Michael Anderson
University of Hartford

Susan Leigh Anderson
University of Connecticut
Contents

General Introduction 1

PART I THE NATURE OF MACHINE ETHICS

Introduction 7
1. The Nature, Importance, and Difficulty of Machine Ethics 13
   James H. Moor
2. Machine Metaethics 21
   Susan Leigh Anderson
3. Ethics for Machines 28
   J. Storrs Hall

PART II THE IMPORTANCE OF MACHINE ETHICS

Introduction 47
   Colin Allen, Wendell Wallach, and Iva Smit
5. Authenticity in the Age of Digital Companions 62
   Sherry Turkle

PART III ISSUES CONCERNING MACHINE ETHICS

Introduction 79
6. What Matters to a Machine? 88
   Drew McDermott
   Steve Torrance
| 8. | On Computable Morality: An Examination of Machines as Moral Advisors | 138 |
|    | Blay Whitby |
| 9. | When Is a Robot a Moral Agent? | 151 |
|    | John P. Sullins |
| 10. | Philosophical Concerns with Machine Ethics | 162 |
|    | Susan Leigh Anderson |
|    | Deborah G. Johnson |
| 12. | On the Morality of Artificial Agents | 184 |
|    | Luciano Floridi |
| 13. | Legal Rights for Machines: Some Fundamental Concepts | 213 |
|    | David J. Calverley |

**PART IV  APPROACHES TO MACHINE ETHICS**

*Introduction*

231

a. **Overview**

14. Towards the Ethical Robot | 244
   James Gips

b. **Asimov’s Laws**

15. Asimov’s Laws of Robotics: Implications for Information Technology | 254
   Roger Clarke

16. The Unacceptability of Asimov’s Three Laws of Robotics as a Basis for Machine Ethics | 285
   Susan Leigh Anderson

c. **Artificial Intelligence Approaches**

17. Computational Models of Ethical Reasoning: Challenges, Initial Steps, and Future Directions | 297
   Bruce M. McLaren

18. Computational Neural Modeling and the Philosophy of Ethics: Reflections on the Particularism-Generalism Debate | 316
   Marcello Guarini
Contents

19. Architectures and Ethics for Robots: Constraint Satisfaction as a Unitary Design Framework  
Alan K. Mackworth  
20. Piagetian Roboethics via Category Theory: Moving beyond Mere Formal Operations to Engineer Robots Whose Decisions Are Guaranteed to be Ethically Correct  
Selmer Bringsjord, Joshua Taylor, Bram van Heuveln, Konstantine Arkoudas, Micah Clark and Ralph Wojtowicz  
21. Ethical Protocols Design  
Matteo Turilli  
22. Modeling Morality with Prospective Logic  
Luís Moniz Pereira and Ari Saptawijaya  

d. Psychological/Sociological Approaches  
23. An Integrated Reasoning Approach to Moral Decision Making  
Morteza Dehghani, Ken Forbus, Emmett Tomai and Matthew Klenk  
Peter Danielson  

e. Philosophical Approaches  
25. There Is No “I” in “Robot”: Robots and Utilitarianism  
Christopher Grau  
26. Prospects for a Kantian Machine  
Thomas M. Powers  
27. A Prima Facie Duty Approach to Machine Ethics: Machine Learning of Features of Ethical Dilemmas, Prima Facie Duties, and Decision Principles through a Dialogue with Ethicists  
Susan Leigh Anderson and Michael Anderson  

PART V  VISIONS FOR MACHINE ETHICS  

Introduction  
28. What Can AI Do for Ethics?  
Helen Seville and Debora G. Field  
29. Ethics for Self-Improving Machines  
J. Storrs Hall
30. How Machines Might Help Us Achieve Breakthroughs in Ethical Theory and Inspire Us to Behave Better  
   Susan Leigh Anderson 524

31. Homo Sapiens 2.0: Building the Better Robots of Our Nature  
   Eric Dietrich 531