

Cambridge University Press

978-0-521-11235-2 - Machine Ethics

Edited by Michael Anderson and Susan Leigh Anderson

Frontmatter

[More information](#)

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.machineethics.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.

Cambridge University Press

978-0-521-11235-2 - Machine Ethics

Edited by Michael Anderson and Susan Leigh Anderson

Frontmatter

[More information](#)

Machine Ethics

Edited by

Michael Anderson

University of Hartford

Susan Leigh Anderson

University of Connecticut



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press

978-0-521-11235-2 - Machine Ethics

Edited by Michael Anderson and Susan Leigh Anderson

Frontmatter

[More information](#)

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521112352

© Cambridge University Press 2011

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2011

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication data
Machine ethics / [edited by] Michael Anderson, Susan Leigh Anderson.
p. cm.

Includes bibliographical references.

ISBN 978-0-521-11235-2 (hardback)

1. Artificial intelligence – Philosophy. 2. Artificial intelligence – Moral and
ethical aspects. I. Anderson, Michael, 1951– II. Anderson, Susan Leigh.

Q335.M165 2011
170–dc22 2010052339

ISBN 978-0-521-11235-2 Hardback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs
for external or third-party Internet Web sites referred to in this publication and does not
guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

Contents

<i>General Introduction</i>	<i>page 1</i>
PART I THE NATURE OF MACHINE ETHICS	
<i>Introduction</i>	7
1. The Nature, Importance, and Difficulty of Machine Ethics James H. Moor	13
2. Machine Metaethics Susan Leigh Anderson	21
3. Ethics for Machines J. Storrs Hall	28
PART II THE IMPORTANCE OF MACHINE ETHICS	
<i>Introduction</i>	47
4. Why Machine Ethics? Colin Allen, Wendell Wallach, and Iva Smit	51
5. Authenticity in the Age of Digital Companions Sherry Turkle	62
PART III ISSUES CONCERNING MACHINE ETHICS	
<i>Introduction</i>	79
6. What Matters to a Machine? Drew McDermott	88
7. Machine Ethics and the Idea of a More-Than-Human Moral World Steve Torrance	115

Cambridge University Press
978-0-521-11235-2 - Machine Ethics
Edited by Michael Anderson and Susan Leigh Anderson
Frontmatter
[More information](#)

vi	Contents	
8.	On Computable Morality: An Examination of Machines as Moral Advisors Blay Whitby	138
9.	When Is a Robot a Moral Agent? John P. Sullins	151
10.	Philosophical Concerns with Machine Ethics Susan Leigh Anderson	162
11.	Computer Systems: Moral Entities but Not Moral Agents Deborah G. Johnson	168
12.	On the Morality of Artificial Agents Luciano Floridi	184
13.	Legal Rights for Machines: Some Fundamental Concepts David J. Calverley	213
	PART IV APPROACHES TO MACHINE ETHICS	
	<i>Introduction</i>	231
	<i>a. Overview</i>	
14.	Towards the Ethical Robot James Gips	244
	<i>b. Asimov's Laws</i>	
15.	Asimov's Laws of Robotics: Implications for Information Technology Roger Clarke	254
16.	The Unacceptability of Asimov's Three Laws of Robotics as a Basis for Machine Ethics Susan Leigh Anderson	285
	<i>c. Artificial Intelligence Approaches</i>	
17.	Computational Models of Ethical Reasoning: Challenges, Initial Steps, and Future Directions Bruce M. McLaren	297
18.	Computational Neural Modeling and the Philosophy of Ethics: Reflections on the Particularism-Generalism Debate Marcello Guarini	316

Cambridge University Press
978-0-521-11235-2 - Machine Ethics
Edited by Michael Anderson and Susan Leigh Anderson
Frontmatter
[More information](#)

Contents	vii
19. Architectures and Ethics for Robots: Constraint Satisfaction as a Unitary Design Framework Alan K. Mackworth	335
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	361
21. Ethical Protocols Design Matteo Turilli	375
22. Modeling Morality with Prospective Logic Luís Moniz Pereira and Ari Saptawijaya <i>d. Psychological/Sociological Approaches</i>	398
23. An Integrated Reasoning Approach to Moral Decision Making Morteza Dehghani, Ken Forbus, Emmett Tomai and Matthew Klenk	422
24. Prototyping N-Reasons: A Computer Mediated Ethics Machine Peter Danielson <i>e. Philosophical Approaches</i>	442
25. There Is No “I” in “Robot”: Robots and Utilitarianism Christopher Grau	451
26. Prospects for a Kantian Machine Thomas M. Powers	464
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	476
PART V VISIONS FOR MACHINE ETHICS	
<i>Introduction</i>	495
28. What Can AI Do for Ethics? Helen Seville and Debora G. Field	499
29. Ethics for Self-Improving Machines J. Storrs Hall	512

Cambridge University Press
978-0-521-11235-2 - Machine Ethics
Edited by Michael Anderson and Susan Leigh Anderson
Frontmatter
[More information](#)

viii	Contents	
30.	How Machines Might Help Us Achieve Breakthroughs in Ethical Theory and Inspire Us to Behave Better Susan Leigh Anderson	524
31.	<i>Homo Sapiens</i> 2.0: Building the Better Robots of Our Nature Eric Dietrich	531