

Contents

<i>Acknowledgments</i>	ix
<i>List of Contributors</i>	xi

A Short Introduction to the Ethics of Artificial Intelligence	1
<i>S. Matthew Liao</i>	

I. BUILDING ETHICS INTO MACHINES

1. Ethical Learning, Natural and Artificial	45
<i>Peter Railton</i>	
2. The Use and Abuse of the Trolley Problem: Self-Driving Cars, Medical Treatments, and the Distribution of Harm	79
<i>Frances M. Kamm</i>	
3. The Moral Psychology of AI and the Ethical Opt-Out Problem	109
<i>Jean-François Bonnefon, Azim Shariff, and Iyad Rahwan</i>	
4. Modeling and Reasoning with Preferences and Ethical Priorities in AI Systems	127
<i>Andrea Loreggia, Nicholas Mattei, Francesca Rossi, and K. Brent Venable</i>	
5. Computational Law, Symbolic Discourse, and the AI Constitution	155
<i>Stephen Wolfram</i>	

II. THE NEAR FUTURE OF ARTIFICIAL INTELLIGENCE

6. Planning for Mass Unemployment: Precautionary Basic Income	183
<i>Aaron James</i>	
7. Autonomous Weapons and the Ethics of Artificial Intelligence	212
<i>Peter Asaro</i>	
8. Near-Term Artificial Intelligence and the Ethical Matrix	237
<i>Cathy O'Neil and Hanna Gunn</i>	
9. The Ethics of the Artificial Lover	271
<i>Kate Devlin</i>	

III. LONG-TERM IMPACT OF SUPERINTELLIGENCE

10. Public Policy and Superintelligent AI: A Vector Field Approach 293
Nick Bostrom, Allan Dafoe, and Carrick Flynn
11. Artificial Intelligence: A Binary Approach 327
Stuart Russell
12. Alignment for Advanced Machine Learning Systems 342
*Jessica Taylor, Eliezer Yudkowsky, Patrick LaVictoire,
and Andrew Critch*
13. Moral Machines: From Value Alignment to Embodied Virtue 383
Wendell Wallach and Shannon Vallor
14. Machines Learning Values 413
Steve Petersen

IV. ARTIFICIAL INTELLIGENCE, CONSCIOUSNESS, AND
MORAL STATUS

15. How to Catch an AI Zombie: Testing for Consciousness
in Machines 439
Susan Schneider
16. Designing AI with Rights, Consciousness, Self-Respect,
and Freedom 459
Eric Schwitzgebel, with Mara Garza
17. The Moral Status and Rights of Artificial Intelligence 480
S. Matthew Liao
- Index* 505