

Paula Boddington

Towards a Code of Ethics for Artificial Intelligence

 Springer

Contents

1	Introduction: Artificial Intelligence and Ethics	1
1.1	Why Ethics in AI? Why Now?	1
1.2	Current Initiatives in AI and Ethics	3
1.3	Codes of Ethics in Context: Other Approaches to Ethical Questions in AI	4
1.3.1	Epistemic Strategies: Precision and the Reduction of Uncertainty	4
1.3.2	Technological Strategies to Ensure Safe and Beneficial AI	5
1.3.3	Moral Strategies in the Pursuit of Beneficial AI	5
2	What Do We Need to Understand About Ethics?	7
2.1	A Preliminary Plea: Ethics Is Not About ‘Banning’ Things	8
2.2	Normative Ethical Theories	8
2.3	Ethics and Empirical Evidence	9
2.4	So Why Do We Even Need Ethics?	9
2.5	So, With What Sort of Issues Is Ethics Concerned?	11
2.6	Who (or What) Is The Proper Object of Moral Concerns, and How Widely Should Our Concerns Extend?	12
2.7	Four Domains of Ethics: Self, Friend, Stranger, World	13
2.8	What Counts as Adequate Justification and Argument in Ethics?	14
2.8.1	How Do We Gain Moral Knowledge?	15
2.8.2	The Elimination of ‘Bias’	16
2.8.3	When Is Ethical Justification ‘Finished’?	17
2.8.4	Can We Necessarily Even Fully Articulate All Our Key Values?	18
2.8.5	Can There Be Such a Thing as Moral Progress?	19
2.9	Moral Relativism, Moral Justification and AI	22
2.10	A Distributed Morality?	23
2.11	Moral Agents	24

2.12	Moral Motivation	24
2.13	AI, Codes of Ethics and the Law	25
3	Does AI Raise Any Distinctive Ethical Questions?	27
3.1	Methodology: Focusing in on Ethical Questions	27
3.1.1	How Do We Identify Ethical Problems as New?	28
3.2	Many Ethical Issues in AI Are Shared with Other Rapidly Developing Technology	28
3.3	Ethical Questions Arise from AI's Typical Use to Enhance, Supplement, or Replace the Work of Humans	29
3.4	We Also Need to Consider the Methods of Production of AI	31
3.5	Hype in AI and Implications for Methodology in Ethics	31
3.5.1	Hype Can Both Distort Our Ethical Reasoning, and Reveal Things of Potential Interest	32
3.5.2	Hype About AI Can Channel Our Thinking About Solutions	33
3.5.3	Impacts of Hype on Moral Thinking	35
3.6	Conclusion	36
4	Codes of Professional Ethics	39
4.1	Introduction: The Varieties of Ethical Codes	39
4.1.1	The Purposes of Codes and Statements of Principle	40
4.2	Professional Codes of Ethics Tend to Have Certain Commonalities	41
4.2.1	Relations Between Professionals, Clients and Others	41
4.2.2	Professional Codes of Ethics, Enforcement, and Authority	42
4.2.3	Professional Codes of Ethics and Professional Values	43
4.2.4	Values Underlying Professional Codes of Ethics	43
4.2.4.1	The Example of Autonomy	44
4.2.4.2	Articulating Values Underlying Professional Codes of Ethics	45
4.2.4.3	Underlying Professional Values May Be Focused Towards Protecting Individuals	47
4.3	Codes of Ethics and Institutional Backing	47
4.4	The Context of Codes of Ethics	48
4.5	Can Codes of Ethics Make the Situation Worse? Yes	53
5	How AI Challenges Professional Ethics	59
5.1	AI Professional Organisations and Companies, and the Nature of Its Development and Production	59
5.2	Gradients of Professional Power and Vulnerability in AI	61
5.3	A Third Layer of Complexity in Codes of Professional Ethics for AI: The Behaviour of Machines	63
5.4	The Authority of Any Resulting Codes	64
5.5	Conclusions	65

6	Developing Codes of Ethics Amidst Fast Technological Change	67
6.1	Social, Cultural and Technological Change and Ethics	67
6.1.1	Methodology and Moral Theory in Times of Change	68
6.1.2	Social and Cultural Change Is Perennial	69
6.1.3	Social, Cultural and Technological Change Is Multifactorial	70
6.1.4	Change and Moral Uncertainty	71
6.2	Social, Cultural, Economic and Technological Change: The Example of AI and Employment	73
6.3	Regulating for Whom? The Global Reach of AI, Universalism, and Relativism	76
6.4	Diversity in Participation as Part of the Solution	79
7	Some Characteristic Pitfalls in Considering the Ethics of AI, and What to Do About Them	85
7.1	The Idealisation of Human and of Machine Agency	85
7.1.1	The Abstract and the Concrete in Ethics	85
7.1.2	Artificial Intelligence, and Intelligence as the Hallmark of Humanity	86
7.1.3	Idealisation and Overreach Often Applies in Thinking About the Ethics of AI	87
7.1.4	Idealisation in Thought About Autonomous Vehicles . . .	88
7.2	Building Ethics into AI and the Idealisation of Moral Agency . . .	89
7.3	Replacing and Enhancing Human Agency, Boundaries and AI . . .	91
7.3.1	Case by Case Consideration Is Needed	91
7.3.2	What Kind of Questions Do We Need to Ask in Such Cases?	92
7.3.3	AI, Ethics, and Effects on Complex Systems	93
7.3.4	Pay Attention: Technology Can Hide, and Technology Can Blind Us	93
7.4	Addressing the Increased Gradient of Vulnerability	94
7.5	Common Language, Miscommunication and the Search for Clarity	95
7.5.1	Common Language May Mask Disagreement: A Tale of Two Autonomies	96
8	Some Suggestions for How to Proceed	99
8.1	Organisations and Codes	99
8.2	Procedures for Drawing Up and Implementing Codes	100
8.3	The Content of Codes	101
8.4	Thinking About Ethical Issues in Developing and Implementing Codes of Ethics	103
8.5	Asilomar AI Principles	104
8.5.1	The Process of Producing the Principles	104
8.5.2	Research Issues in the Asilomar Principles	105

8.5.3	Ethics and Values in the Asilomar Principles	106
8.5.4	Longer-Term Issues in AI	110
8.5.5	General Comments on the Asilomar Principles	110
References	113
Index	121