

# Contents

<i>Preface</i>	xi
<i>Acknowledgements</i>	xix
<i>List of Figures</i>	xxiii
<i>List of Tables</i>	xxv
<i>List of Most Common Abbreviations and Acronyms</i>	xxvii

## PART ONE UNDERSTANDING AI

1. Past: The Emergence of AI	3
1.0 Summary	3
1.1 Introduction: The Digital Revolution and AI	3
1.2 Digital's Cleaving Power: Cutting and Pasting Modernity	6
1.3 New Forms of Agency	10
1.4 AI: A Research Area in Search of a Definition	12
1.5 Conclusion: Ethics, Governance, and Design	12
2. Present: AI as a New Form of Agency, Not Intelligence	14
2.0 Summary	14
2.1 Introduction: What Is AI? 'I Know It When I See It'	14
2.2 AI as a Counterfactual	16
2.3 The Two Souls of AI: The Engineering and the Cognitive	20
2.4 AI: A Successful Divorce Made in the Infosphere	24
2.5 The Human Use of Humans as Interfaces	27
2.6 Conclusion: Who Will Adapt to Whom?	29
3. Future: The Foreseeable Development of AI	31
3.0 Summary	31
3.1 Introduction: Looking into the Seeds of Time	31
3.2 Historical, Hybrid, and Synthetic Data	32
3.3 Constraining and Constitutive Rules	37
3.4 Difficult Problems, Complex Problems, and the Need for Enveloping	39
3.5 Generative Models	43
3.6 A Future of Design	49
3.7 Conclusion: AI and Its Seasons	50

## PART TWO EVALUATING AI

4. A Unified Framework of Ethical Principles for AI	57
4.0 Summary	57

4.1	Introduction: Too Many Principles?	58
4.2	A Unified Framework of Five Principles for Ethical AI	58
4.3	Beneficence: Promoting Well-Being, Preserving Dignity, and Sustaining the Planet	61
4.4	Nonmaleficence: Privacy, Security, and 'Capability Caution'	61
4.5	Autonomy: The Power to 'Decide to Decide'	62
4.6	Justice: Promoting Prosperity, Preserving Solidarity, Avoiding Unfairness	62
4.7	Explicability: Enabling the Other Principles through Intelligibility and Accountability	63
4.8	A Synoptic View	64
4.9	AI Ethics: Whence and for Whom?	64
4.10	Conclusion: From Principles to Practices	65
5.	From Principles to Practices: The Risks of Being Unethical	67
5.0	Summary	67
5.1	Introduction: Risky Translations	67
5.2	Ethics Shopping	68
5.3	Ethics Bluewashing	70
5.4	Ethics Lobbying	71
5.5	Ethics Dumping	72
5.6	Ethics Shirking	74
5.7	Conclusion: The Importance of Knowing Better	76
6.	Soft Ethics and the Governance of AI	77
6.0	Summary	77
6.1	Introduction: From Digital Innovation to the Governance of the Digital	77
6.2	Ethics, Regulation, and Governance	79
6.3	Compliance: Necessary but Insufficient	81
6.4	Hard and Soft Ethics	82
6.5	Soft Ethics as an Ethical Framework	84
6.6	Ethical Impact Analysis	87
6.7	Digital Preferability and the Normative Cascade	88
6.8	Digital Ethics' Dual Advantage	89
6.9	Conclusion: Ethics as a Strategy	90
7.	Mapping the Ethics of Algorithms	92
7.0	Summary	92
7.1	Introduction: A Working Definition of Algorithm	92
7.2	Map of the Ethics of Algorithms	94
7.3	Inconclusive Evidence Leading to Unjustified Actions	96
7.4	Inscrutable Evidence Leading to Opacity	97
7.5	Misguided Evidence Leading to Unwanted Bias	101
7.6	Unfair Outcomes Leading to Discrimination	103

7.7	Transformative Effects Leading to Challenges for Autonomy and Informational Privacy	106
7.8	Traceability Leading to Moral Responsibility	109
7.9	Conclusion: The Good and Evil Use of Algorithms	111
8.	Bad Practices: The Misuse of AI for Social Evil	113
8.0	Summary	113
8.1	Introduction: The Criminal Use of AI	113
8.2	Concerns	115
8.2.1	Emergence	117
8.2.2	Liability	117
8.2.3	Monitoring	119
8.2.4	Psychology	120
8.3	Threats	121
8.3.1	Commerce, Financial Markets, and Insolvency	121
8.3.2	Harmful or Dangerous Drugs	123
8.3.3	Offences against the Person	124
8.3.4	Sexual Offences	127
8.3.5	Theft and Fraud, Forgery, and Personation	129
8.4	Possible Solutions	131
8.4.1	Tackling Emergence	131
8.4.2	Addressing Liability	133
8.4.3	Checking the Monitoring	135
8.4.4	Dealing with the Psychology	137
8.5	Future Developments	138
8.5.1	Areas of AIC	138
8.5.2	Dual-Use	138
8.5.3	Security	139
8.5.4	Persons	139
8.5.5	Organizations	140
8.6	Conclusion: From Evil Uses of AI to Socially Good AI	140
9.	Good Practices: The Proper Use of AI for Social Good	142
9.0	Summary	142
9.1	Introduction: The Idea of AI for Social Good	142
9.2	A Definition of AI4SG	146
9.3	Seven Essential Factors for Successful AI4SG	148
9.3.1	Falsifiability and Incremental Deployment	149
9.3.2	Safeguards against the Manipulation of Predictors	151
9.3.3	Receiver-Contextualized Intervention	153
9.3.4	Receiver-Contextualized Explanation and Transparent Purposes	155
9.3.5	Privacy Protection and Data Subject Consent	159
9.3.6	Situational Fairness	161
9.3.7	Human-Friendly Semanticization	163
9.4	Conclusion: Balancing Factors for AI4SG	164