CONTENTS

Acknowledgments *xi*

Introduction **1**

Chapter 1: What Is Artificial Intelligence and Why Does It Raise Philosophically Interesting Ethical Issues? **6**

1: The Blake Lemoine Incident 6

2: What Is Artificial Intelligence? An Evolving Concept 9

2.1: Thinking Machines? Turing vs. Jefferson 10

2.2. The Invention of the Term "Artificial Intelligence" and the Question of Whether AI Technologies Are Intelligent 13

3: What Definition of AI Should We Accept? Exclusive vs. Inclusive Approaches 17

4: Narrow and Broad Conceptions of—and Negative and Positive Approaches to—the Ethics of Artificial Intelligence 19

5: Why Does the Notion of Artificial Intelligence Raise Philosophically Interesting Ethical Questions? 21

6: Outlook 23

Chapter 2: Control and Value Alignment Problems, AI Hype, and the Notion of Artificial General Intelligence **27**

1: The 2022 Release of ChatGPT and the AI Hype and Worries that Followed 27

2: What Are the Control and Value Alignment Problems? And What Does "Artificial General Intelligence" Refer To? 31

3: Is There Really a Control Problem and, If So, (How) Could It Be Solved? 34

4: Human-Compatible AI and Superalignment 37

5: An Out-of-Control Hype that Distracts Our Attention Away from More Pressing Ethical Problems Associated with Contemporary AI? 42

6: Do We Have to Make a Choice Between *Either* Focusing on Risks Related to Future AGI *or* Focusing on Harms and Risks Related to Today's AI Technologies? 46

7: Revisiting the Question of Whether AI Technologies Possess Any Forms of Agency and Intelligence 49

Chapter 3: Responsibility and AI, Good and Bad Outcomes, and Asymmetries Between Praise and Blame 57

1: Lee Sedol's Loss Against AlphaGo and the First Time a Self-Driving Car Killed a Human Being 57

2: A General Introduction to the Idea of Moral Responsibility 61

3: Ten Important Points About Moral Responsibility and About Praise and Blame and the Differences Between Them 64

4: Artificial Intelligence and Responsibility Gaps 70

5: Filling Responsibility Gaps by Having People Volunteer to Take Responsibility? 74

6: Praiseworthiness and Blameworthiness for Good and Bad Outcomes Created by/with AI Technologies 77

7: Concluding Remarks 81

Chapter 4: Generative AI's Gappiness with Respect to Meaning, Authorship, and Credit and Blame 84

1: From Hollywood Screenwriters to Poisonous Mushrooms 84

2: Gappiness? 87

3: Reminder: Credit-Blame Asymmetries 88

4: Meaning, Understanding, and Worries About a Lack of Meaning and Understanding in Generative AI Technologies 90

5: The Chinese Room and the Octopus 94

6: The Death of the Author, the Return of the Author, and the Question of Whether Generative AI Technologies Can Be Authors 98

7: Authorship Gaps: Can Generative AI Be an Author or Co-Author? 102

8: Should Human Users Always Be Considered the Authors of the Outputs of Generative AI Technologies? 106

9: Do We Need New Norms, Concepts, and Theories to Deal with Generative AI's Gappiness? 107

Chapter 5: Artificial Intelligence and Human Enhancement: Do AI Technologies Make Us More or Less (Artificially) Intelligent? 112

1: Back to the Man Who Was Carrying Out the Winning Moves in the AlphaGo Games Against Lee Sedol 112

2: Human Intelligence and Cognitive Enhancement 115

3: Artificial Intelligence and Its Relation to Human Intelligence 118

4: Might AI Technologies Enhance Human Intelligence by Extending Our Minds? 120

5: Artificial Human Intelligence 125

6: Concluding Discussion 129

Chapter 6: Artificial Intelligence, Consciousness, and the Moral Status of Conscious and Non-Conscious AI Agents 138

1: People in the Tech World Who Claim that AI Technologies Are Conscious 138

2: The Basic Concepts in This Chapter 141

3: A Quick Way of Mapping Some of the Relevant Recent Literature About AI Consciousness (or Lack Thereof) and Moral Status (or Lack Thereof) 147

4: Are Any AI Agents Conscious and/or Sentient or Might They Soon Be? 149

5: Consciousness and Moral Status 153

6: AI Agents Without Consciousness/Sentience but with Moral Status? 156

7: Critical Discussion of the Views Presented in the Previous Section 159

Chapter 7: Personalized Imitation Games: Permissibility, Desirability, and the Ethics of Digital Duplicates and AI Copies of Real People 168

1: From AI-Generated Pop Stars to Digital Duplicates of Dead Relatives 168

2: What Is a Personalized Digital Duplicate/AI Imitation of a Particular Person? 171

3: Personalized Imitation Games and AI Replacement Worries 174

4: The Permissibility of Personalized Digital Duplicates 178

5: The Desirability of Personalized Digital Duplicates 184

6: Conclusions of this Chapter and Some Further Discussion of Digital Duplicates 189

Chapter 8: Moral Principles for Human-Human Interaction, Human-AI Interaction, and AI-AI Interaction 195

1: Should AI Technologies Follow the Same Moral Principles as (Particular) Human Beings? 195

2: Should AI Technologies Follow Moral Principles in the First Place? 198

3: Moral Agents and Moral Patients, and Three Kinds of Interaction: Human-Human, Human-AI, and AI-AI Interaction 202

4: Traditional Ethical Principles and Principles for the Ethics of AI 208

5: The Double Standards Argument 211

6: Possible Justifications for Both Double and Triple Standards 214

7: Concluding Remarks and Transition to the Next Chapter 218

Chapter 9: The Future of (Meaningful) Work, Meaning in Life, and Human Flourishing in the Age of AI **223**

1: Benefits for All of Humanity? 223

2: Different Kinds of Improvements to Human Lives 225

3: Can AI Make Human Lives Morally Better? 228

4: Five Quick Arguments for Why AI Can Have an Impact on Meaning in Life 231

5: The Achievement Gap and the Future of (Meaningful) Work 235

6: The Prospects for Meaning and Human Flourishing Within Futuristic AI Utopias 240

7: The AI and Meaning Sweet Spot 246

Chapter 10: AI Risks and Opportunities, Sustainability, and the Ethics of Gambling with the Future of Humanity **252**

1: The AI Wager 252

2: The Ethics of Risk Imposition 255

3: AI and Different Kinds of Sustainability 260

4: The Drunk Driver Analogy and the "Who Are the Moral Agents and Who Are the Moral Patients?" Problem 264

5: Utopian Optimism vs. Dystopian Pessimism, and Responsible and Reckless Types of AI Optimism 268

6: A Young, Developing, and Quickly Growing Field 270

Concluding Summary: Retracing Our Steps **275**

Bibliography *280*

Index *294*