

CONTENTS

<i>Lists of Figures, Tables, and Boxes</i>	xv
1. Past developments and present capabilities	1
Growth modes and big history	1
Great expectations	4
Seasons of hope and despair	6
State of the art	14
Opinions about the future of machine intelligence	22
2. Paths to superintelligence	26
Artificial intelligence	27
Whole brain emulation	35
Biological cognition	43
Brain–computer interfaces	54
Networks and organizations	58
Summary	61
3. Forms of superintelligence	63
Speed superintelligence	64
Collective superintelligence	65
Quality superintelligence	68
Direct and indirect reach	70
Sources of advantage for digital intelligence	71
4. The kinetics of an intelligence explosion	75
Timing and speed of the takeoff	75
Recalcitrance	80
<i>Non-machine intelligence paths</i>	80
<i>Emulation and AI paths</i>	82
Optimization power and explosivity	89
5. Decisive strategic advantage	95
Will the frontrunner get a decisive strategic advantage?	96

How large will the successful project be?	101
<i>Monitoring</i>	102
<i>International collaboration</i>	104
From decisive strategic advantage to singleton	106
6. Cognitive superpowers	110
Functionalities and superpowers	111
An AI takeover scenario	115
Power over nature and agents	120
7. The superintelligent will	127
The relation between intelligence and motivation	127
Instrumental convergence	131
<i>Self-preservation</i>	132
<i>Goal-content integrity</i>	132
<i>Cognitive enhancement</i>	134
<i>Technological perfection</i>	136
<i>Resource acquisition</i>	137
8. Is the default outcome doom?	140
Existential catastrophe as the default outcome of an intelligence explosion?	140
The treacherous turn	142
Malignant failure modes	146
<i>Perverse instantiation</i>	146
<i>Infrastructure profusion</i>	149
<i>Mind crime</i>	153
9. The control problem	155
Two agency problems	155
Capability control methods	157
<i>Boxing methods</i>	158
<i>Incentive methods</i>	160
<i>Stunting</i>	163
<i>Tripwires</i>	167
Motivation selection methods	169
<i>Direct specification</i>	169
<i>Domesticity</i>	172
<i>Indirect normativity</i>	173
<i>Augmentation</i>	173
Synopsis	175

10. Oracles, genies, sovereigns, tools	177
Oracles	177
Genies and sovereigns	181
Tool-AIs	184
Comparison	190
11. Multipolar scenarios	194
Of horses and men	195
<i>Wages and unemployment</i>	195
<i>Capital and welfare</i>	197
<i>The Malthusian principle in a historical perspective</i>	199
<i>Population growth and investment</i>	201
Life in an algorithmic economy	203
<i>Voluntary slavery, casual death</i>	204
<i>Would maximally efficient work be fun?</i>	207
<i>Unconscious outsourcers?</i>	210
<i>Evolution is not necessarily up</i>	212
Post-transition formation of a singleton?	216
<i>A second transition</i>	216
<i>Superorganisms and scale economies</i>	218
<i>Unification by treaty</i>	220
12. Acquiring values	226
The value-loading problem	226
Evolutionary selection	229
Reinforcement learning	230
Associative value accretion	231
Motivational scaffolding	233
Value learning	235
Emulation modulation	246
Institution design	247
Synopsis	253
13. Choosing the criteria for choosing	256
The need for indirect normativity	256
Coherent extrapolated volition	259
<i>Some explications</i>	260
<i>Rationales for CEV</i>	262
<i>Further remarks</i>	264
Morality models	266
Do What I Mean	270

Component list	271
<i>Goal content</i>	272
<i>Decision theory</i>	274
<i>Epistemology</i>	275
<i>Ratification</i>	277
Getting close enough	278
14. The strategic picture	280
Science and technology strategy	281
<i>Differential technological development</i>	281
<i>Preferred order of arrival</i>	283
<i>Rates of change and cognitive enhancement</i>	286
<i>Technology couplings</i>	291
<i>Second-guessing</i>	293
Pathways and enablers	295
<i>Effects of hardware progress</i>	295
<i>Should whole brain emulation research be promoted?</i>	297
<i>The person-affecting perspective favors speed</i>	302
Collaboration	303
<i>The race dynamic and its perils</i>	303
<i>On the benefits of collaboration</i>	306
<i>Working together</i>	311
15. Crunch time	314
Philosophy with a deadline	314
What is to be done?	315
<i>Seeking the strategic light</i>	317
<i>Building good capacity</i>	317
<i>Particular measures</i>	318
Will the best in human nature please stand up	319
Afterword	321
<i>Notes</i>	325
<i>Bibliography</i>	383
<i>Partial Glossary</i>	407
<i>Index</i>	411