

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	3
Seasons of hope and despair	5
State of the art	11
Opinions about the future of machine intelligence	18
2. Paths to superintelligence	22
Artificial intelligence	23
Whole brain emulation	30
Biological cognition	36
Brain-computer interfaces	44
Networks and organizations	48
Summary	50
3. Forms of superintelligence	52
Speed superintelligence	53
Collective superintelligence	54
Quality superintelligence	56
Direct and indirect reach	58
Sources of advantage for digital intelligence	59
4. The kinetics of an intelligence explosion	62
Timing and speed of the takeoff	62
Recalcitrance	66
<i>Non-machine intelligence paths</i>	66
<i>Emulation and AI paths</i>	68
Optimization power and explosivity	73

5. Decisive strategic advantage	78
Will the frontrunner get a decisive strategic advantage?	79
How large will the successful project be?	83
<i>Monitoring</i>	84
<i>International collaboration</i>	86
From decisive strategic advantage to singleton	87
6. Cognitive superpowers	91
Functionalities and superpowers	92
An AI takeover scenario	95
Power over nature and agents	99
7. The superintelligent will	105
The relation between intelligence and motivation	105
Instrumental convergence	109
<i>Self-preservation</i>	109
<i>Goal-content integrity</i>	109
<i>Cognitive enhancement</i>	111
<i>Technological perfection</i>	112
<i>Resource acquisition</i>	113
8. Is the default outcome doom?	115
Existential catastrophe as the default outcome of an intelligence explosion?	115
The treacherous turn	116
Malignant failure modes	119
<i>Perverse instantiation</i>	120
<i>Infrastructure profusion</i>	122
<i>Mind crime</i>	125
9. The control problem	127
Two agency problems	127
Capability control methods	129
<i>Boxing methods</i>	129
<i>Incentive methods</i>	131
<i>Stunting</i>	135
<i>Tripwires</i>	137
Motivation selection methods	138
<i>Direct specification</i>	139
<i>Domesticity</i>	140
<i>Indirect normativity</i>	141
<i>Augmentation</i>	142
Synopsis	143

10. Oracles, genies, sovereigns, tools	145
Oracles	145
Genies and sovereigns	148
Tool-AIs	151
Comparison	155
11. Multipolar scenarios	159
Of horses and men	160
<i>Wages and unemployment</i>	160
<i>Capital and welfare</i>	161
<i>The Malthusian principle in a historical perspective</i>	163
<i>Population growth and investment</i>	164
Life in an algorithmic economy	166
<i>Voluntary slavery, casual death</i>	167
<i>Would maximally efficient work be fun?</i>	169
<i>Unconscious outsourcers?</i>	172
<i>Evolution is not necessarily up</i>	173
Post-transition formation of a singleton?	176
<i>A second transition</i>	177
<i>Superorganisms and scale economies</i>	178
<i>Unification by treaty</i>	180
12. Acquiring values	185
The value-loading problem	185
Evolutionary selection	187
Reinforcement learning	188
Associative value accretion	189
Motivational scaffolding	191
Value learning	192
Emulation modulation	201
Institution design	202
Synopsis	207
13. Choosing the criteria for choosing	209
The need for indirect normativity	209
Coherent extrapolated volition	211
<i>Some explications</i>	212
<i>Rationales for CEV</i>	213
<i>Further remarks</i>	216
Morality models	217
Do What I Mean	220
Component list	221
<i>Goal content</i>	222

<i>Decision theory</i>	223
<i>Epistemology</i>	224
<i>Ratification</i>	225
Getting close enough	227
14. The strategic picture	228
Science and technology strategy	228
<i>Differential technological development</i>	229
<i>Preferred order of arrival</i>	230
<i>Rates of change and cognitive enhancement</i>	233
<i>Technology couplings</i>	236
<i>Second-guessing</i>	238
Pathways and enablers	240
<i>Effects of hardware progress</i>	240
<i>Should whole brain emulation research be promoted?</i>	242
<i>The person-affecting perspective favors speed</i>	245
Collaboration	246
<i>The race dynamic and its perils</i>	246
<i>On the benefits of collaboration</i>	249
<i>Working together</i>	253
15. Crunch time	255
Philosophy with a deadline	255
What is to be done?	256
<i>Seeking the strategic light</i>	257
<i>Building good capacity</i>	258
<i>Particular measures</i>	258
Will the best in human nature please stand up	259
<i>Notes</i>	261
<i>Bibliography</i>	305
<i>Index</i>	325