A READER'S GUIDE TO NICK BOSTROM'S

SUPERINTELLIGENCE
Paths, Dangers, Strategies

MIRI

VERSION 1

1 HOW TO USE THIS GUIDE

Nick Bostrom's *Superintelligence: Paths, Dangers, Strategies* (2014) is a meaty work, and it is best digested one bite at a time. This reader's guide breaks the book into 30 manageable sections, usually of 6-10 pages each. (See part 2.)

This guide can be used by individuals or groups who are reading through the book at any pace, but it was originally envisioned as a guide for online reading groups, or for in-person reading groups that meet weekly to discuss the material together. Reading is most fun and enlightening when done with others.

Ambitious readers, or those who have been following FHI's and MIRI's research for a year or more, may wish to cover two sections per week. All other readers are advised to digest one section at a time.

One could also try to read most of the book during a single semester of school, which is usually 12-15 weeks in the USA. (See part 3.)

This guide is an early version. Readers are encouraged to fill out this feedback form to suggest any additions or modification to future guides.

Discussion questions recommended for every sections include:

- What did you find confusing or unclear?
- What did you find most persuasive?
- What did you find least persuasive?
- What do you most want to learn more about?
- What did you change your mind about as a result of the reading?

2 HOW TO READ THE BOOK IN 30 SECTIONS

Section 01: **Past developments and present capabilities**
Reading: Foreword & "Growth modes…" through "State of the art" from Chapter 1

Section 02: **Forecasting AI**
Reading: "Opinions about the future of machine intelligence" from Chapter 1
Muehlhauser, "When Will AI Be Created?"

Section 03: **AI & whole brain emulation**
Reading: "Artificial intelligence" and "Whole brain emulation" from Chapter 2

Section 04: **Biological cognition, BCIs, organizations**
Reading: "Biological cognition" and the rest of Chapter 2

Section 05: **Forms of superintelligence**
Reading: Chapter 3

Section 06: **Intelligence explosion kinetics**
Reading: Chapter 4

Section 07: **Decisive strategic advantage**
Reading: Chapter 5

Section 08: **Cognitive superpowers**
Reading: Chapter 6

Section 09: **The orthogonality of intelligence and goals**
Reading: "The relation between intelligence and motivation" from Chapter 7

Section 10: **Instrumentally convergent goals**
Reading: "Instrumental convergence" from Chapter 7

Section 11: **The treacherous turn**
Reading: "Existential catastrophe…" and "The treacherous turn" from Chapter 8

Section 12: **Malignant failure modes**
Reading: "Malignant failure modes" from Chapter 8

Section 13: **Capability control methods**
Reading: "Two agency problems" and "Capability control methods" from Chapter 9

Section 14: **Motivation selection methods**
Reading: "Motivation selection methods" and "Synopsis" from Chapter 9

Section 15: **Oracles, genies, and sovereigns**
Reading: "Oracles" and "Genies and Sovereigns" from Chapter 10

[Continued on following page]

Section 16: **Tool AIs**
Reading: "Tool-AIs" and "Comparison" from Chapter 10

Section 17: **Multipolar scenarios**
Reading: "Of horses and men" from Chapter 11

Section 18: **Life in an algorithmic economy**
Reading: "Life in an algorithmic economy" from Chapter 11

Section 19: **Post-transition formation of a singleton**
Reading: "Post-transition formation of a singleton?" from Chapter 11

Section 20: **The value-loading problem**
Reading: "The value-loading problem" through "Motivational scaffolding" from Chapter 12

Section 21: **Value learning**
Reading: "Value learning" from Chapter 12

Section 22: **Emulation modulation and institutional design**
Reading: "Emulation modulation" through "Synopsis" from Chapter 12

Section 23: **Coherent extrapolation volition**
Reading: "The need for..." and "Coherent extrapolated volition" from Chapter 13

Section 24: **Morality models and "Do what I mean"**
Reading: "Morality models" and "Do what I mean" from Chapter 13

Section 25: **Component list for acquiring values**
Reading: "Component list" and "Getting close enough" from Chapter 13

Section 26: **Science and technology strategy**
Reading: "Science and technology strategy" from Chapter 14

Section 27: **Pathways and enablers**
Reading: "Pathways and enablers" from Chapter 14

Section 28: **Collaboration**
Reading: "Collaboration" from Chapter 14

Section 29: **Crunch time**
Reading: Chapter 15

Section 30: [TBD]
Reading: [TBD]

3 HOW TO READ MOST OF THE BOOK IN 12 AMBITIOUS WEEKS

Week 1: Read the Foreword, and also read "Artificial Intelligence", "Whole Brain Emulation," and "Summary" from Chapter 2.

Week 2: Read Chapter 3.

Week 3: Read Chapter 4.

Week 4: Read Chapter 5.

Week 5: Read Chapter 6.

Week 6: Read Chapter 7.

Week 7: Read Chapter 8.

Week 8: Read "Two Agency Problems" and "Capability Control Methods" from Chapter 9.

Week 9: Read "Motivation Selection Methods" and "Synopsis" from Chapter 9.

Week 10: Read Chapter 11.

Week 11: Read "Science and Technology Strategy" from Chapter 14.

Week 12: Read "Pathways and Enablers" and "Collaboration" from Chapter 14, and also read Chapter 15.