Superintelligence: Paths, Dangers, Strategies
Superintelligence asks the questions: What happens when machines surpass humans in general intelligence? Will artificial agents save or destroy us? Nick Bostrom lays the foundation for understanding the future of humanity and intelligent life. The human brain has some capabilities that the brains of other animals lack. It is to these distinctive capabilities that our species owes its dominant position. If machine brains surpassed human brains in general intelligence, then this new superintelligence could become extremely powerful - possibly beyond our control. As the fate of the gorillas now depends more on humans than on the species itself, so would the fate of humankind depend on the actions of the machine superintelligence. But we have one advantage: We get to make the first move. Will it be possible to construct a seed Artificial Intelligence, to engineer initial conditions so as to make an intelligence explosion survivable? How could one achieve a controlled detonation? This profoundly ambitious and original book breaks down a vast track of difficult intellectual terrain. After an utterly engrossing journey that takes us to the frontiers of thinking about the human condition and the future of intelligent life, we find in Nick Bostrom’s work nothing less than a reconceptualization of the essential task of our time.

Book Information
Audible Audio Edition
Listening Length: 14 hours and 16 minutes
Program Type: Audiobook
Version: Unabridged
Publisher: Audible Studios
Audible.com Release Date: September 3, 2014
Language: English
ASIN: B00LPMFE9Y
Best Sellers Rank:
#4 in Books > Computers & Technology > Computer Science > AI & Machine Learning > Intelligence & Semantics
#12 in Books > Audible Audiobooks > Nonfiction > Computers

Customer Reviews
Prof. Bostrom has written a book that I believe will become a classic within that subarea of Artificial Intelligence (AI) concerned with the existential dangers that could threaten humanity as the result of the development of artificial forms of intelligence. What fascinated me is that Bostrom has approached the existential danger of AI from a perspective that, although I am an AI professor, I had
never really examined in any detail. When I was a graduate student in the early 80s, studying for my PhD in AI, I came upon comments made in the 1960s (by AI leaders such as Marvin Minsky and John McCarthy) in which they mused that, if an artificially intelligent entity could improve its own design, then that improved version could generate an even better design, and so on, resulting in a kind of “chain-reaction explosion” of ever-increasing intelligence, until this entity would have achieved “superintelligence”. This chain-reaction problem is the one that Bostrom focusses on. He sees three main paths to superintelligence: 1. The AI path -- In this path, all current (and future) AI technologies, such as machine learning, Bayesian networks, artificial neural networks, evolutionary programming, etc. are applied to bring about a superintelligence. 2. The Whole Brain Emulation path -- Imagine that you are near death. You agree to have your brain frozen and then cut into millions of thin slices. Banks of computer-controlled lasers are then used to reconstruct your connectome (i.e., how each neuron is linked to other neurons, along with the microscopic structure of each neuron’s synapses). This data structure (of neural connectivity) is then downloaded onto a computer that controls a synthetic body.

Not surprisingly, 200+ pages later, the author can’t answer the ‘what is to be done’ question concerning the likely emergence of non-human (machine-based) super-intelligence, sometime, possibly soon. This is expected because, as a species, we’ve always been the smartest ones around and never had to even think about the possibility of coexistence alongside something or someone impossibly smart and smart in ways well beyond our comprehension, possibly driven by goals we can’t understand and acting in ways that may cause our extinction. Building his arguments on available data and extrapolating from there, Bostrom is confident that:- some form of self-aware, machine super-intelligence is likely to emerge- we may be unable to stop it, even if we wanted to, no matter how hard we tried- while we may be unable to stop the emergence of super-intelligence, we could prepare ourselves to manage it and possibly survive it- us not taking this seriously and not being prepared may result in our extinction while serious pre-emergence debate and preparation may result in some form of co-existenceIt’s radical and perhaps frightening but our failure to comprehend the magnitude of the risks we are about to confront would be a grave error given that, once super-intelligence begins to manifest itself and act, the change may be extremely quick and we may not be afforded a second chance. Most of the book concerns itself with the several types of super-intelligence that may develop, the ways in which we may be able to control or at least co-exist with such entities or entity, what the world and literally the Universe may turn into depending on how we plant the initial super-intelligent seed.
There are three pieces to this book: 1. Superhuman machine intelligence is coming. 2. It's potentially catastrophic for humanity. 3. We might be able to tame it. "Superintelligence" is not human-level artificial intelligence as in meeting the Turing Test. It's what comes after that. Once we build a machine as smart as a human, that machine writes software to improve itself, which enables it to further improve itself -- but faster, then faster and faster. The equal-to-human stage proves brief as the technology charges ahead into superhuman territory. Before long, the thing is way smarter than any human, and it parleys its superhuman skill at programming into superhuman skill at everything else, including strategic thinking, practical psychology and propaganda. It can start innocuously enough as a device that only thinks, but using its superhuman strategic skills it persuades humans to give it control over physical resources like manipulators and nanofactories. At that point it becomes a deadly threat even without "wanting" to. For example, its goal could be perfecting battery technology, but as it pursues this goal it could decide it needs more computing resources (to really figure things out...) -- a lot more resources, and so it proceeds to turn half the earth’s crust into computing machinery, while tiling the rest of the planet with solar cells to power that machinery! We lose. If you’re scared now, then you’ll want to read the second 50% or so of the book, which is about how to tame superintelligence, or for now, how to prepare to tame it before it’s already upon us and out of control.

Download to continue reading...

Dmca