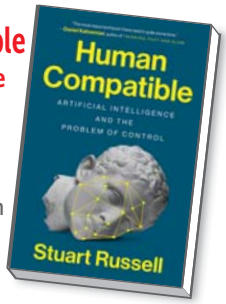
GADO IMAGES/ALAMY STOCK PHOTO

Can artificial superintelligence match its hype?

Human Compatible
Artificial Intelligence and the Problem of Control
Stuart Russell
Viking/Penguin Random House, 2019. $28.00



At a recent meeting of the World Economic Forum, someone asked Stuart Russell, a professor of computer science at the University of California, Berkeley, when superintelligent artificial intelligence (AI) might arrive. He loosely estimated it to be within his children's lifetime, and then he emphasized the Chatham House rules of the meeting and that his conjecture was "strictly off the record." But, he writes in his new book *Human Compatible: Artificial Intelligence and the Problem of Control*, "Less than two hours later, an article appeared in the *Daily Telegraph* citing Professor Russell's remarks, complete with images of rampaging Terminator robots."

Hyperbole by many media outlets has made it challenging for experts to talk seriously about the dangers of artificial superintelligence—a technology that would surpass the intellectual capabilities of humans. Nonetheless, many experts have written books on the subject. Nick Bostrom's 2014 book *Superintelligence: Paths, Dangers, Strategies* raised eyebrows for its passages on embryo selection and modification as a potential path to superintelligence. Murray Shanahan provided only a short introduction to the field in his 2015 book *The Technological Singularity*. And Max Tegmark's 2017 book *Life 3.0: Being Human in the Age of Artificial Intelligence* focuses mostly on ways in which an artificial superintelligence might be horrible to us.

Human Compatible has a more practical and down-to-earth approach, if one can say that about a book on superintelligent AI. Calmly taking the fearmongering headlines in stride, Russell starts at the beginning with a thorough explanation of intelligence, what AI is and does, and how we might reach superintelligence. Then he explains how humanity can and should make sure that its eventual arrival will be beneficial for humanity.

Russell has been one of the foremost academics in the field of AI since the late 1980s, and many would say that a popular book on AI by him is long overdue, especially considering his decades of public advocacy. He is known for two major achievements. First, Russell pioneered inverse-reinforcement learning in 1998, which he explains clearly in *Human Compatible* without trumpeting his own achievements. Second, he has educated generations of AI researchers with *Artificial Intelligence: A Modern Approach*, the textbook he coauthored with Peter Norvig in 1995; the fourth edition is due in 2020.

For those seeking a more accessible introduction to AI, *Human Compatible* provides one of the clearest explanations of the underlying concepts. The scope of the book underscores the vast, interdisciplinary field. Russell explains concepts from computer science, robotics, psychology, economics, mathematics, and politics. Readers without a degree in AI can follow his descriptions, although his definition of intelligence might cause head-scratching: Russell claims that "machines are *beneficial* to the extent that *their* actions can be expected to achieve *our* objectives." It is a confusing utilitarian view that translates better to computer programming than to human behavior.

Russell's excellent writing is, at times, surprisingly funny and sets his work apart from that of his peers. He has a knack for eminently quotable turns of phrase. The chapter on AI misuses begins with a warning of "the rapid rate of innovation in the malfeasance sector." The elegant writing highlights my main contention: Too many people are writing about the dangers of superintelligence rather than more pressing issues such as algorithmic injustice.

In a later chapter, "The Not-So-Great AI Debate," Russell debunks the arguments against taking the risk of superintelligence seriously. He starts with an easy one—calculators and horses haven't taken over the world, so we don't have anything to fear from superhuman intel-

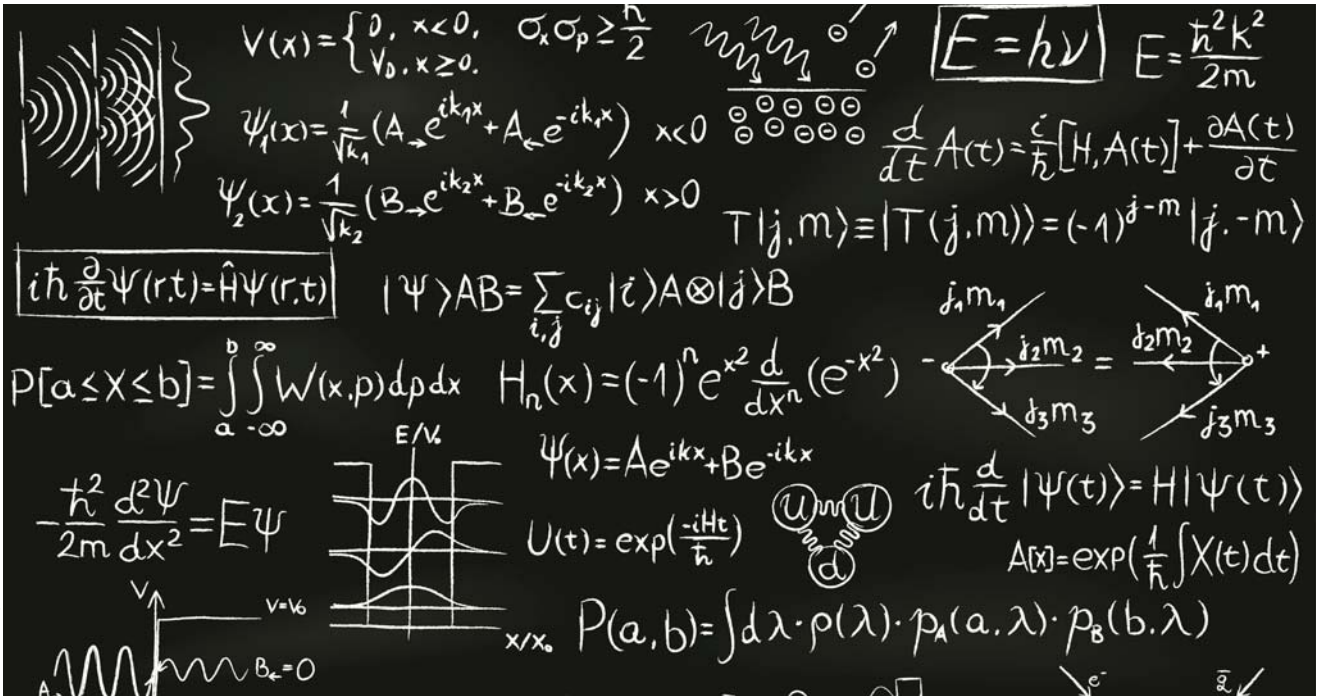
ligence. He moves on to refuting more sophisticated arguments, such as the assertion that AI won't have destructive emotional traits if we don't build them into it. He concludes the chapter with a quote by *Slate Star Codex* blogger Scott Alexander: "We should probably get a couple of bright people to start working on preliminary aspects of the problem."

Notably, that chapter is one of the few times that Russell addresses the beneficial uses of AI. Whereas his treatment of contemporary AI ethics issues is commendable, he unfortunately pays little attention to the technologies that are beginning to show promising results. For example, short-term weather forecasts and long-term climate change projections are both improving because of AI technologies that can crunch vast amounts of data.

I would have loved to see Russell take on a question central to the current "not-so-great AI debate," as he calls it: Should we be paying so much attention to superintelligence when humanity is creating enough problems for itself with existing AI? The relentless growth of contemporary AI technologies, such as driverless cars that require electricity-hungry computer servers to store and process data, threatens the climate, our physical safety, and our privacy. Achieving superintelligence within our children's lifetime poses a significantly lower risk than the possibility that they will be rubbing sticks together for fire after surviving catastrophic global warming or a world war.

Russell concludes that researchers should study superintelligence but that focusing too much attention on it may leave other threats in the AI sector understudied and underfunded. Yes, superintelligence skeptics and activists alike would agree that a few brilliant people should think about superintelligence, but this skeptic thinks the emphasis should fall on "a few."

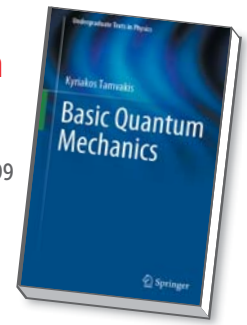
Kanta Dihal
University of Cambridge
Cambridge, UK



ISTOCK.COM/OKTALSTUDIO

Quantum mechanics textbook teaches through examples

Basic Quantum Mechanics
Kyriakos Tamvakis
Springer, 2019. $74.99 (paper)



At the turn of the 20th century, experimentalists began uncovering mysterious phenomena that were unexplained by classical theories. Physicists put forward groundbreaking new ideas, the consequences of which we are still trying to fully comprehend. The modern version of quantum mechanics was developed in the 1920s through pioneering works by Erwin Schrödinger, Werner Heisenberg, Max Born, and other contemporaries. Since then quantum mechanics has become an integral part of standard academic curricula for university physics, and several canonical textbooks exist on the subject.

Quantum mechanics has gained a reputation for being a difficult subject due in part to both its conceptual differences from classical physics and its difficult mathematical machinery. To deal with those challenges, most students learn about quantum mechanics from