

NATIONAL BESTSELLER

WHEN COMPUTERS EXCEED
HUMAN INTELLIGENCE

THE AGE OF SPIRITUAL MACHINES



RAY KURZWEIL
AUTHOR OF *THE AGE OF INTELLIGENT MACHINES*

"The Age of Spiritual Machines will blow your mind. Kurzweil lays out a scenario that might seem like science fiction if it weren't coming from a proven entrepreneur."

— *San Francisco Chronicle*

PROLOGUE: AN INEXORABLE EMERGENCE

The gambler had not expected to be here. But on reflection, he thought he had shown some kindness in his time. And this place was even more beautiful and satisfying than he had imagined. Everywhere there were magnificent crystal chandeliers, the finest handmade carpets, the most sumptuous foods, and, yes, the most beautiful women, who seemed intrigued with their new heaven mate. He tried his hand at roulette, and amazingly his number came up time after time. He tried the gaming tables, and his luck was nothing short of remarkable: He won game after game. Indeed his winnings were causing quite a stir, attracting much excitement from the attentive staff, and from the beautiful women.

This continued day after day, week after week, with the gambler winning every game, accumulating bigger and bigger earnings. Everything was going his way. He just kept on winning. And week after week, month after month, the gambler's streak of success remained unbreakable.

After a while, this started to get tedious. The gambler was getting restless; the winning was starting to lose its meaning. Yet nothing changed. He just kept on winning every game, until one day, the now anguished gambler turned to the angel who seemed to be in charge and said that he couldn't take it anymore. Heaven was not for him after all. He had figured he was destined for the "other place" nonetheless, and indeed that is where he wanted to be.

"But this is the other place," came the reply.

That is my recollection of an episode of *The Twilight Zone* that I saw as a young child. I don't recall the title, but I would call it "Be Careful What You Wish For." As this engaging series was wont to do, it illustrated one of the paradoxes of human nature: We like to solve problems, but we don't want them all solved, not too quickly, anyway. We are more attached to the problems than to the solutions.

Take death, for example. A great deal of our effort goes into avoiding it. We make extraordinary efforts to delay it, and indeed often consider its intrusion a tragic event. Yet we would find it hard to live without it. Death gives meaning to our lives. It gives importance and value to time. Time would become meaningless if there were too much of it. If death were indefinitely put off, the human psyche would end up, well, like the gambler in *The Twilight Zone* episode.

We do not yet have this predicament. We have no shortage today of either death or human problems. Few observers feel that the twentieth century has left us with too much of a good thing. There is growing prosperity, fueled not incidentally by information technology, but the human species is still challenged by issues and difficulties not altogether different than those with which it has struggled from the beginning of its recorded history.

The twenty-first century will be different. The human species, along with the computational technology it created, will be able to solve age-old problems of need, if not desire, and will be in a position to change the nature of mortality in a postbiological future. Do we have the psychological capacity for all the good things that await us? Probably not. That, however, might change as well.

Before the next century is over, human beings will no longer be the most intelligent or capable type of entity on the planet. Actually, let me take that back. The truth of that last statement depends on how we define human. And here we see one profound difference between these two centuries: The primary political and philosophical issue of the next century will be the definition of who we are.²

But I am getting ahead of myself. This last century has seen enormous technological change and the social upheavals that go along with it, which few pundits circa 1899 foresaw. The pace of change is accelerating and has been since the inception of invention (as I will discuss in the first chapter, this acceleration is an inherent feature of technology). The result will be far greater transformations in the first two decades of the twenty-first century than we saw in the entire twentieth century. However, to appreciate the inexorable logic of where the twenty-first century will bring us, we have to go back and start with the present.

TRANSITION TO THE TWENTY-FIRST CENTURY

Computers today exceed human intelligence in a broad variety of intelligent yet narrow domains such as playing chess, diagnosing certain medical conditions, buying and selling stocks, and guiding cruise missiles. Yet human intelligence overall remains far more supple and flexible. Computers are still unable to describe the objects on a crowded kitchen table, write a summary of a movie, tie a pair of shoelaces, tell the difference between a dog and a cat (although this feat, I

believe, is becoming feasible today with contemporary neural nets—computer simulations of human neurons),³ recognize humor, or perform other subtle tasks in which their human creators excel.

One reason for this disparity in capabilities is that our most advanced computers are still simpler than the human brain—currently about a million times simpler (give or take one or two orders of magnitude depending on the assumptions used). But this disparity will not remain the case as we go through the early part of the next century. Computers doubled in speed every three years at the beginning of the twentieth century, every two years in the 1950s and 1960s, and are now doubling in speed every twelve months. This trend will continue, with computers achieving the memory capacity and computing speed of the human brain by around the year 2020.

Achieving the basic complexity and capacity of the human brain will not automatically result in computers matching the flexibility of human intelligence. The organization and content of these resources—the software of intelligence—is equally important. One approach to emulating the brain's software is through reverse engineering—scanning a human brain (which will be achievable early in the next century)⁴ and essentially copying its neural circuitry in a neural computer (a computer designed to simulate a massive number of human neurons) of sufficient capacity.

There is a plethora of credible scenarios for achieving human-level intelligence in a machine. We will be able to evolve and train a system combining massively parallel neural nets with other paradigms to understand language and model knowledge, including the ability to read and understand written documents. Although the ability of today's computers to extract and learn knowledge from natural-language documents is quite limited, their abilities in this domain are improving rapidly. Computers will be able to read on their own, understanding and modeling what they have read, by the second decade of the twenty-first century. We can then have our computers read all of the world's literature—books, magazines, scientific journals, and other available material. Ultimately, the machines will gather knowledge on their own by venturing into the physical world, drawing from the full spectrum of media and information services, and sharing knowledge with each other (which machines can do far more easily than their human creators).

Once a computer achieves a human level of intelligence, it will necessarily roar past it. Since their inception, computers have significantly exceeded human mental dexterity in their ability to remember and process information. A computer can remember billions or even trillions of facts perfectly, while we are hard pressed to remember a handful of phone numbers. A computer can quickly search a database with billions of records in fractions of a second. Computers

can readily share their knowledge bases. The combination of human-level intelligence in a machine with a computer's inherent superiority in the speed, accuracy, and sharing ability of its memory will be formidable.

Mammalian neurons are marvelous creations, but we wouldn't build them the same way. Much of their complexity is devoted to supporting their own life processes, not to their information-handling abilities. Furthermore, neurons are extremely slow; electronic circuits are at least a million times faster. Once a computer achieves a human level of ability in understanding abstract concepts, recognizing patterns, and other attributes of human intelligence, it will be able to apply this ability to a knowledge base of all human-acquired—and machine-acquired—knowledge.

A common reaction to the proposition that computers will seriously compete with human intelligence is to dismiss this specter based primarily on an examination of contemporary capability. After all, when I interact with my personal computer, its intelligence seems limited and brittle, if it appears intelligent at all. It is hard to imagine one's personal computer having a sense of humor, holding an opinion, or displaying any of the other endearing qualities of human thought.

But the state of the art in computer technology is anything but static. Computer capabilities are emerging today that were considered impossible one or two decades ago. Examples include the ability to transcribe accurately normal continuous human speech, to understand and respond intelligently to natural language, to recognize patterns in medical procedures such as electrocardiograms and blood tests with an accuracy rivaling that of human physicians, and, of course, to play chess at a world-championship level. In the next decade, we will see translating telephones that provide real-time speech translation from one human language to another, intelligent computerized personal assistants that can converse and rapidly search and understand the world's knowledge bases, and a profusion of other machines with increasingly broad and flexible intelligence.

In the second decade of the next century, it will become increasingly difficult to draw any clear distinction between the capabilities of human and machine intelligence. The advantages of computer intelligence in terms of speed, accuracy, and capacity will be clear. The advantages of human intelligence, on the other hand, will become increasingly difficult to distinguish.

The skills of computer software are already better than many people realize. It is frequently my experience that when demonstrating recent advances in, say, speech or character recognition, observers are surprised at the state of the art. For example, a typical computer user's last experience with speech-recognition technology may have been a low-end freely bundled piece of software from several years ago that recognized a limited vocabulary, required pauses between words, and did an incorrect job at that. These users are then surprised to see

contemporary systems that can recognize fully continuous speech on a 60,000-word vocabulary, with accuracy levels comparable to a human typist.

Also keep in mind that the progression of computer intelligence will sneak up on us. As just one example, consider Gary Kasparov's confidence in 1990 that a computer would never come close to defeating him. After all, he had played the best computers, and their chess-playing ability—compared to his—was pathetic. But computer chess playing made steady progress, gaining forty-five rating points each year. In 1997, a computer sailed past Kasparov, at least in chess. There has been a great deal of commentary that other human endeavors are far more difficult to emulate than chess playing. This is true. In many areas—the ability to write a book on computers, for example—computers are still pathetic. But as computers continue to gain in capacity at an exponential rate, we will have the same experience in these other areas that Kasparov had in chess. Over the next several decades, machine competence will rival—and ultimately surpass—any particular human skill one cares to cite, including our marvelous ability to place our ideas in a broad diversity of contexts.

Evolution has been seen as a billion-year drama that led inexorably to its grandest creation: human intelligence. The emergence in the early twenty-first century of a new form of intelligence on Earth that can compete with, and ultimately significantly exceed, human intelligence will be a development of greater import than any of the events that have shaped human history. It will be no less important than the creation of the intelligence that created it, and will have profound implications for all aspects of human endeavor, including the nature of work, human learning, government, warfare, the arts, and our concept of ourselves.

This specter is not yet here. But with the emergence of computers that truly rival and exceed the human brain in complexity will come a corresponding ability of machines to understand and respond to abstractions and subtleties. Human beings appear to be complex in part because of our competing internal goals. Values and emotions represent goals that often conflict with each other, and are an unavoidable by-product of the levels of abstraction that we deal with as human beings. As computers achieve a comparable—and greater—level of complexity, and as they are increasingly derived at least in part from models of human intelligence, they, too, will necessarily utilize goals with implicit values and emotions, although not necessarily the same values and emotions that humans exhibit.

A variety of philosophical issues will emerge. Are computers thinking, or are they just calculating? Conversely, are human beings thinking, or are they just calculating? The human brain presumably follows the laws of physics, so it must be a machine, albeit a very complex one. Is there an inherent difference between

human thinking and machine thinking? To pose the question another way, once computers are as complex as the human brain, and can match the human brain in subtlety and complexity of thought, are we to consider them conscious? This is a difficult question even to pose, and some philosophers believe it is not a meaningful question; others believe it is the only meaningful question in philosophy. This question actually goes back to Plato's time, but with the emergence of machines that genuinely appear to possess volition and emotion, the issue will become increasingly compelling.

For example, if a person scans his brain through a noninvasive scanning technology of the twenty-first century (such as an advanced magnetic resonance imaging), and downloads his mind to his personal computer, is the "person" who emerges in the machine the same consciousness as the person who was scanned? That "person" may convincingly implore you that "he" grew up in Brooklyn, went to college in Massachusetts, walked into a scanner here, and woke up in the machine there. The original person who was scanned, on the other hand, will acknowledge that the person in the machine does indeed appear to share his history, knowledge, memory, and personality, but is otherwise an impostor, a different person.

Even if we limit our discussion to computers that are not directly derived from a particular human brain, they will increasingly appear to have their own personalities, evidencing reactions that we can only label as emotions and articulating their own goals and purposes. They will appear to have their own free will. They will claim to have spiritual experiences. And people—those still using carbon-based neurons or otherwise—will believe them.

One often reads predictions of the next several decades discussing a variety of demographic, economic, and political trends that largely ignore the revolutionary impact of machines with their own opinions and agendas. Yet we need to reflect on the implications of the gradual, yet inevitable, emergence of true competition to the full range of human thought in order to comprehend the world that lies ahead.

PART ONE

PROBING THE PAST