Since the industrial age fired up its first machines to replace human labor, our collective imagination has been haunted by the shadow of the robot.

The Shadow of the Robot
Since the industrial age fired up its first machines to replace human labor, our collective imagination has been haunted by the shadow of the robot, the artificial intelligence (AI) that so disturbingly mimics our own consciousness, encased within a metal exoskeleton. From the clunky tin cans of early Amazing Stories-style sci-fi to the modern image of a sleek android that is physically, mentally and intellectually indistinguishable from us—indeed often superior to us—these human-shaped thinking machines have always played deeply into our anxieties about what it means to be human. Yet, at the same time, they have suggested some great liberation for the human race, perhaps even some form of immortality. The dichotomy between a desired end—the need for technology to free humanity from toil—and the price we pay for attempting to achieve those ends—our growing dependency on the machinery, perhaps at the expense of our essential humanity—creates vast anxiety.

This anxiety has been an undercurrent in our literary culture for almost two centuries, since William Blake first poeticized against the “dark Satanic mills” that were turning England’s pastoral landscape into an enslaving industrial engine. The genre of science fiction, which evolved over the decades along with technology itself, has always reflected this anxiety by exploring...
our dark thoughts and fearful dreams relating to the tyranny of our own machines. At the same time, science fiction has been a literary celebration of technological progress, as if that very progress brings us closer to some liberating event that can only be achieved through the evolution of our machines. Both the drive to progress via technology and the fear of that very same technology as some dehumanizing force have been at the heart of our science fiction literature almost since its inception.

In the film Matrix Reloaded, Neo takes a tour of the vast underground city of humans that had been built as a refuge from the machines that have aggressively conquered man. Gazing at the colossal engines that run the city, he instinctively wants to believe that these machines that are bringing him light and heat and food are different from the killer machines—dominating the planet because these are machines that can be turned off at our will. The irony of that moment is that the humans simply cannot shut off the life-bringing machines even if they wanted to. For one thing, no one alive even remembers how they work. Second, what would happen to their daily life if they did? In a moment’s pause, Neo wrestles with the notion that the existential experience of being human may be based on how dependent we are on the machinery that we have built (Silver, 2003).

What would happen to a nation like the United States if we lost the computers that delivered our electricity, if we lost the chips that ran the automobiles and the transportation systems, or the circuits that produced the food and purified the waters? Even if after a period of readjustment (no doubt one filled with unimaginable violence and suffering) we managed to return to an industrial agrarian society, similar to the one we had a hundred years ago, how many people would die in the interim? And once such a replacement society was established, how would we feed, house and keep warm the vast population that we have today?

Is it true that the machines and computers have permeated so deep into our lives that we are no longer independent of them, that we are now committed to this exo-skeletal nervous system we call technology? It is no surprise that our current anxieties about our machines increasingly center on the notion that we cannot live without them, but that perhaps, in a science fiction kind of way, they will one day “wake up” and decide that they indeed can live without us?

Science Fiction and the Revolt of the Machines
In April of 2000, Bill Joy, the co-founder of Sun Microsystems, published an alarming article in Wired magazine entitled “Why the Future Doesn’t Need Us” (Joy, 2000, p.238). He had just met Ray Kurzweil, an accomplished inventor whose musical synthesizer had changed the face of popular music; Kurzweil was also (and remains) a pioneer in the field of speech recognition technologies and other AI initiatives. Kurzweil had recently published a book, The Age of Spiritual Machines (1999), and after engaging the author in an animated discussion about the book’s darker implications, Bill Joy was deeply disturbed.

The Age of Spiritual Machines is a sweeping vision of how computer technology is expected to evolve over the next few decades, vividly showing how Moore’s Law—the tendency of data density on a computer chip to double every 18 months—predicts that at some point in the near future, a single computer chip will be more powerful than a human brain. In the book, Kurzweil speaks of self-replicating nanobots, human minds downloaded like software into android bodies, and the fleshy human body itself slowly being infiltrated (Kurzweil, 1999), like the Borg of Star Trek: The Next Generation, with machine implants fuzzing the lines between humans and computers.

The machines we humans have built to extend our senses and increase our own intellect’s productivity, Kurzweil says, will soon be “waking up” to consciousness and will begin the construction of even more advanced machinery. Robot factories will not only build better and more powerful robots, but design them as well, not always with human interests in mind. By the year 2099, Kurzweil concludes, there will be no clear distinction between humans and computers. Intelligent machines will by then become “spiritual machines,” meaning that they will legally be recognized as a species in their own right and have legitimate consciousness equal to, if not surpassing, our own (Kurzweil, 1999).

While this all seems like something out of a Hollywood movie to be enjoyed with a bucket of popcorn, Kurzweil builds a very convincing argument that it is at least technologically possible, and that the computer industry of the last 60 years was a mere prelude in this vast shift in consciousness that will follow the emergence of these spiritual machines. This is the point that made Bill Joy very upset and prompted him to write his article for Wired (Joy, 2000). Joy had spent his career taking pleasure in being part of a new paradigm; he’d watched computer science evolve from the geeky hacking culture that created the first operating systems, programming languages and cheaper, more powerful computer chips into the big industry of today that drives the global economy and runs the machinery that supports our daily lives.

“But while I was aware of the moral dilemmas surrounding technology’s consequences in fields like weapons research,” Joy wrote in his article, “I did not expect that I would confront such issues in my own field, or at least not so soon” (Joy, 2000). Ray Kurzweil seemed calm and collected about the dawn of the spiritual machines and the surrendering of much of what we consider to be our humanity, but Joy was walking about in a panic saying, “Does anyone else here think this is dangerous and dehumanizing?” In fact, he would have to look no further than the literary
genre of science fiction—a genre that, for more than a century, has been openly exploring the anxieties relating to the surrender of our humanity to our own machines—in order to find people thoughtful enough to agree with him.

Such depictions often reflected the largest fear of that age: the revolt of the worker. Almost from its inception, science fiction has been relegated to the ghetto of pulp magazines and reduced by Hollywood into action-adventure space operas, always struggling to maintain its literary respectability and promote understanding of its social import. The genre found its biggest audience in adolescent males and was quickly laughed off as “Buck Rogers stuff”: its marketing niche was on the same bookshelves as dime store detectives, gun-toting cowboys and masked super heroes.

Yet, from the beginning, the genre attracted many highly intelligent and visionary writers who weren’t as interested in the swashbuckling adventures of planet hopping Errol Flynn-types (as in the popular stories of Edgar Rice Burroughs), but in the potential and possibilities inherent in the rapid exponential growth of technology and all the ethical, moral and philosophical quandaries that surround such growth. In the 1920s, when writers like Jack Williamson, Murray Leinster, Edmond Hamilton and E.E. “Doc” Smith1 were pioneering new directions in storytelling, the world about them was changing at a breakneck pace. A dynamic expansion of technology made us optimistic that the human race could be improved through technology; but at the same time we suffered tremendous anxiety over the potential misuse of the very machines that were improving our lives.

For this reason, early science fiction showed both an enthusiasm and a dread of new technologies. The classical image of the mad alchemist who, to paraphrase director Ed Wood, tamps in God’s domain (Wood, 1955) and pays the price for his trespasses, portrayed to perfection in Mary Shelley’s Frankenstein (2003), was transformed, by the early 1900s, into the materialistic scientist, who was so obsessed by his new invention (be it an invisibility ray, a time machine or a way to harness nuclear power), that he could not see its physical dangers or moral implications until it was too late. Many sci-fi classics such as Isaac Asimov’s I, Robot (1950) or Robert Heinlein’s The Moon Is a Harsh Mistress (1966) have dealt with the ethical dilemmas posed by the growing field of robotics and the potential for AI. But none of the technologies described in any of this literature had actually been invented yet. The sheer fact that we can imagine such machines taking over our lives is enough to fire the creative juices and makes for great drama. The best science fiction writers were like canaries in the coalmine, small creatures whose early death signaled to the workers that the air in the mine shafts had turned rank and deadly and that it was time to evacuate or suffer annihilation.

Science fiction has always been a Freudian journey through our collective anxieties, arrayed in the trappings of entertainment. Contemporary political situations like the Red Scare and the Cold War, as well as controversial technologies like the atom bomb and the computer, have all fed the story mills of the sci-fi imagination. While many sci-fi writers appeared on the surface to be just “spinning a good yarn,” even the Star Wars-style space operas played upon our fears of evil empires and our growing dehumanization, as typified by Darth Vader, the man who has lost his humanity to robotic body parts.

H.G. Wells’ The War of the Worlds (1898) depicted a Martian race that appears amorphous and blob-like, hardly a threat to our military until they construct their tripods and heat rays that blow away human civilization like it is a cluster of insects. Wells’ novel heralded the violence of World War One, where tanks and airplanes appeared for the first time on the battlefields and showed the world how the horror of the Martian Invasion was possible within our lifetime, and may be brought about by our own machinery. Depictions of robots in early science fiction reflected fears of rapid industrialization, showing them as clanky humanoid hunks of metal filled with gears and wires. Such depictions often reflected the largest fear of that age: the revolt of the worker. Industrial capitalists were witnessing the social power of communism and other revolutionary movements that threatened to empower the workers who were enslaved in their factories, revealing that they were actual human beings with needs, feelings and, perhaps worse, the resources to organize resistance. The deepest source of this drive towards revolution was the manner in which capitalist industrialism dehumanized the worker, threatened to turn him into a cog in a machine (a robot, if you will) and, without doubt, sci-fi literature dealt with the consequences of that dehumanization. Early masterpieces like the play R.U.R. (Rossum’s Universal Robots) (Cape, 2004) and the motion picture Metropolis (Pommer, 1927) did overly carry the robot-as-worker metaphor into the political arena and showed audiences the dangers of mechanizing life itself in the name of progress and profit.

Over the decades, the source of the various anxieties that gave formative shape to the sci-fi plots and dilemmas changed in variance with historical, political and social conditions. The “evil empire” sci-fi of Flash Gordon paralleled the rise of Fascism in Europe; the adventures in the deserts of Mars or the jungles of Venus drew from the growing exploration into the Congo and the Amazon where white Europeans came in contact with cultures living in Paleolithic conditions; and the subduing of savage alien tribes by a Caucasian Earthling in many tales by Edgar Rice Burroughs validated American Manifest Destiny2 and the destruction of Native American Indian culture. During the 1950s, the menace of the Pod People in The Body Snatchers (Finney, 1989) played upon fears of the communist infiltration of our own society, or in preference to a non-political interpretation, the xenophobic fear of the outsider invading our communities. No longer was the menace a force from another planet, or another alien or robotic race; now the danger was inside of us, biding its time inside our own neighbors and spouses. Then the horrors of radiation—an invisible force more sinister than any communist country, unleashed by the military and their nuclear hardware—gave rise to giant monsters and mutations.

Finally, almost inevitably, came the threat of AI, the computer that considers itself superior to the human race. In the mountain-sized mainframe of the Colossus from Dennis Feltham Jones’ 1967 novel of the same name and HAL9000 of 2001: A Space Odyssey (Kubrick, 1968), we have villains to rival (if not exceed) any of Flash Gordon’s antagonists. In both stories, the computers built by man to monitor

1 The character of Anthony “Buck” Rogers, a US Air Corps officer who awakens from a coma in the 25th Century, was one of the first heroes of the science fiction pulp. He first appeared in the August 1928 issue of Amazing Stories and was the creation of Philip Francis Nolan. Many of the subsequent sci-fi heroes were modeled after Buck, including Flash Gordon and Luke Skywalker, but the character also stood as the stereotype of silly pulp melodrama. The iconography of sci-fi including ray guns and space ships became known as “Buck Roger’s stuff” and was not always used as a compliment.

2 Many of these authors published prolifically in pulp magazines such as Amazing Stories, Astounding Science Fiction, Weird Tales and The Magazine of Fantasy & Science Fiction and their early works can be found in many good anthologies that draw from these magazines. Some of the better science fiction pieces, such as "Triplanetary," by E.E. "Doc" Smith and "The Legion of Time" by Jack Williamson have become classics of the genre.

3 Edgar Rice Burroughs created several characters who were the epitome of the heroic individualist transferred to another planet. John Carter of Mars, Carson Napier of Venus and David Innes of Pellucidar (the world at the Earth’s core) were all white Americans who conquer an alien race living in Paleolithic conditions, usually becoming either King, Emperor or Warlord, despite the fact that of their being a racial outsider. A fringe benefit to the job was usually a marriage to the most beautiful woman on the planet. One cannot help but see a reflection of the domination of the Native Americans by white Europeans in the conquering of Mars by John Carter. Burroughs’ most famous creation, Tarzan of the Apes, was another example of this formula.
itor and control our national security and life support systems suffer acute paranoia and lash back at humans like they were unwanted factors in a new world order.

But still, the computer was exterior to our selves, bland voices emerging from voice units, cold, calculating circuitry that couldn’t possibly feel or relate subjectively to the world. Our advantage over them was their inability to feel or think like humans. All we had to do to release ourselves from the tyranny of these machines was pull their plug and watch them go as lifeless as a television tuned to a dead channel.

But what would happen, as in the case of HAL9000 (Kubrick, 1968), if the computer knew it was being turned off—suddenly cared about its existence, to our food supply, to our heat and air—suddenly felt the urge to create just as well as we can, if not better?

What do we do when confronted with a species that can work and think and create just as well as we can, if not better?

Philip K. Dick and the Empathy Factor

While the problem of how humans respond to the growing dominance of the thinking machines is an old theme in science fiction dating back to its origins, the ethical complexities and moral conundrums relating to it grew more sophisticated as the genre grew into its third and fourth generation of writers. One of the keystone novels of the modern period of sci-fi that dealt with this issue was Do Androids Dream of Electric Sheep? (1968), by Philip K. Dick, a prolific writer who not only explored the social and political anxieties relating to technology, but extended those anxieties into consciousness itself. Dick, who suffered many personal neuroses and phobias (among which was the frightening feeling that he didn’t exist), wrote many haunted novels which have had a long-term influence on the genre. His dark footprints can be seen in films such as The Matrix (Silver, 1999), The Truman Show (Rudin, 1998) and Videodrome (Heroux, 1983), all of which feature a main character who, after having his sense of reality severely ruptured, struggles to determine what is real and what is fabricated by some inhuman machine with its own agenda.

In more than 30 novels, both sci-fi and mainstream, and over 100 short stories, Dick explored the possible disintegration of his own consciousness and his desperation in trying to define what is human. In Do Androids Dream of Electric Sheep?, he presents his clearest answer to the question, “What does it mean to be human?” (Dick, 1968). Written during a decade of feverish activity and coming on the heels of many potboilers that Dick wrote to pay the rent, it is a stunning blend of commercial science fiction, existential and Gnostic philosophy, and the exploration of the author’s personal demons. While there are many Dick novels that are better written, this is one of his most endearing and most memorable.

The dystopic world of Do Androids Dream of Electric Sheep? (Dick, 1968) is typical of many Philip K. Dick novels throughout the 1960s, one in which a global war has radiated the cities and killed off most animal life—“First, strangely, the owls had died.” (Dick, p.12)—giving rise to a robotics industry that manufactures artificial pets. A mad rush to colonize Mars has left the earth underpopulated, largely by genetically inferior people who are left to wallow in the slowly building garbage and useless rubble that advances in the cow catcher of entropy. To fulfill the daily needs of the Martian colonists, the robotics industry has also produced the Nexus-6, a brand of android that is almost indistinguishable from human beings, but which can labor without complaint and give the colonists an unprecedented degree of leisure. The Nexus-6 are illegal on Earth, and an elite corps of bounty hunters within the police department have been trained to detect and terminate any androids that have escaped to Earth.

Rick Deckard, a bounty hunter disgruntled with his life and career, is very much a Philip Dickian Everyman, a wage laborer hustling to get contracts so he can raise the money to buy a real animal to replace the mechanical sheep that he covets on his rooftop garden. As the novel opens, Deckard’s liberal wife is chastising him for being an assassin, which he denies, since the androids he terminates are not human in the first place, and therefore his work is not murder. In short, he does not recognize the essential selfhood of the androids and rationalizes that the major difference between humans and androids is that the “andys” do not have the ability to empathize with the suffering of other creatures. In fact, the only way to legally prove that someone is an android, short of testing their bone marrow, is to apply a complex psycho-physical examination that tests their neuromuscular reaction to tales of animal suffering.

Here, again, is the essential existential problem: what is it that makes us human? Here, again, is the essential existential problem: what is it that makes us human? We can easily distinguish ourselves from the lower life forms and even the apes, but what do we do when confronted with a species that can work and think and create just as well as we can, if not better? Who is to say that we will still maintain our status as the dominant species? When the androids have outmoded us and proven to be superior, and everything that we coveted as being part of our humanity is being done better by another species, what remains that makes us essentially human?

Philip K. Dick’s answer is Empathy.

In many of his novels, Dick professed a fondness for the soulful presence of animals and was greatly discomforted by people who had no empathy for an animal’s suffering. This lack of empathy was one step removed from the same soullessness that drove the Nazis to dehumanize their enemies and kill them by the millions with heartless efficiency. To Dick, the difference between humans and androids must be the ability to feel empathy for the suffering of others.

In the novel, the humans who have stayed behind on Earth, some because they could not afford to emigrate, others because they have been deemed inferior by the government, are struggling to maintain their humanity. It is implied that those who have emigrated to Mars have given up their humanity, delegating their labor and drudgery to the androids, and surrendered the remote possibility of clean air and healthy landscapes for artificial machine-driven ecosystems. Back on Earth, the remaining...
humans, fighting against the onslaught of entropy and fearful of the superior race of outlawed androids, have adapted a quasi-religious practice called Mercerism to keep in touch with their gradually diminishing humanity.

Mercerism is practiced by the followers of Wilbur Mercer, who may or may not exist. He can only be “accessed” by an empathy box, a virtual reality generator that plugs anyone gripping its handles into a weird Gogolgotha-like landscape where a robed and bearded figure, Wilbur Mercer, ascends a steep mountain slope. From out of visual range, hostile “others” are hurling rocks at Mercer, and the person plugged into the empathy box feels the pain as if the rock has hit him personally. Mercer’s climb up the hill is experienced, both visually and physically, by everyone who is currently using the Mercer machine, as well as those who are locked together in an empathic experience of shared suffering.

While the Christian overtones here are more than obvious, we can dismiss Dick’s use of Mercerism as primarily Christian. The meditation upon the wounds of Christ is a well-known meditational technique to stimulate empathic abilities and to awaken one’s consciousness to the subjective suffering of others, but other religions and psycho-spiritual practices, such as Tibetan Buddhism, place an emphasis on compassion for all sentient life forms as well. Whether Mercer exists or not is irrelevant: what matters it that he is a quasi-religious practice called Mercerism, which shattered their loneliness and makes them feel connected. It is assumed that androids cannot have that experience since the empathy box of Wilbur Mercer has no effect on them.

Deckard is at first skeptical of his wife’s use of the Mercer machine, and Mercerism, seen by many to be a quack religion, is under assault by a gross and obnoxious TV comedian named Buster Friendly (one in a series of ongoing Dick characters who resemble the 1950s celebrity Jackie Gleason) who threatens to expose Mercer as a fake, an actor in a studio. The suspicion that Buster Friendly is a pawn of the androids—who have an acute self-interest in destroying Mercerism—is heightened by the discovery that the androids have set up an elaborate operation on Earth, going as far as to engage their own underground police force to capture the bounty hunters. The android’s will to survive has grown effectively stronger and many of them are exhibiting the human-like emotions of empathy for each other.

In fact, as Rick Deckard goes about hunting down the androids so he can collect his bounty money and buy a real flesh and blood animal, he confronts the intelligence and the apparent inner aesthetic world of the androids. Nexus-6 are the latest and greatest, more human than human, and not even the corporation who built them understands the inner dynamics of their own subjective world.

Deckard confronts one of the aliens while she is headlining an opera event, and he is seduced by the powerful music and the android’s very human singing of Mozart’s The Magic Flute. Later, in an art museum, he contemplates why so many androids are drawn to the works of Edvard Munch, The Scream in particular, which he assumes is a painting that reflects the inner feelings of an android. Predictably, in a very painful turnabout, Deckard starts feeling contempt and indifference towards the other bounty hunters and acquires more respect for the machines he is contracted to destroy. In fact, he makes physical love to one of the androids in order to render himself incapable of killing her.

The use of the opera and the art museum in the novel is very significant, because as humans we like to think that we are the only species that can produce a Mozart or a Munch. But in The Age of Spiritual Machines, Kurzweil (1999) describes AI machines that will at some point become so sophisticated that they will paint pictures, write poetry, solve ethical problems, evaluate law and compose music so much better than us. In the light of this usurpation of our essential aesthetic humanity, the existential dilemma of what it means to be human will become vastly more acute.

We insist that, no matter how good a computer gets at composing music, it was ultimately Beethoven’s spiritual life that made his symphonies great, something a computer can never achieve. So Kurzweil goes to great pains to present us with examples of art created by computers—not some imaginary spiritual computer, but computer programs that exist today. He includes some paintings, poems and even a short story. At first blush, it seems incredible that a program can produce works with such aesthetic impact, but Kurzweil reveals that the programs relied on being fed input from existing paintings, poetry and fiction that had been created under the more traditional inspiration of human desire and intuition. Without the artistic work of human beings, the computer-generated art would not exist.

There is a pulsing light at the heart of artistic creation that these computer outputs seem to lack. The question is whether or not, at some point in the future when AI programs become more powerful than the human brain, that pulsing light will suddenly erupt and awaken the machines to their new consciousness. Will computers then be capable of writing a poem simply out of the sheer joy and love of the language, or to express some inner state of being, or be capable of producing a complex harmony of aesthetics such as is exhibited in pieces like Beethoven’s Ninth Symphony?

Perhaps the spiritual outcries we hear in the Ninth will also be heard in a piece of music composed by some android two hundred years from now and come from a complexity of forces of consciousness that were not possible when created by the human mind alone. Perhaps the computers, besides being stronger, more intelligent and less fragile than us, will also be more spiritual than us. Perhaps the new emerging spiritual machines will stand in relation to us as we do to animals and look upon us as artifacts of an earlier stage of evolution.

The Egyptian Immortality Game
All this talk of inputs and generating poems through analysis seems so spiritually enlightened, like one of those episodes of Star Trek: The Next Generation in which Data the Android struggles to understand the concept of finding something funny. However, by meditating upon the science fiction conceit of a machine becoming human, or a human becoming machine, we are forced to confront the subjective experience of being a human consciousness with a phenomenological interrelationship with an organic body.
Kurzweil explores the possibility that we will gradually merge with our machines and become some form of hybrid race, that our culture, our flesh, our minds will gradually be replaced, one circuit at a time, by computer components, and that the very concept of a computer itself will be redefined by incorporating our very humanity and our very flesh into a new species. According to Kurzweil, this is the logical continuation of an evolutionary process that goes back all the way to the Big Bang, as if the creation of sub-atomic particles out of the primordial Chaos was the Universe’s first step towards the creation of speech recognition software and really cool operating systems. It can be left to the individual to decide just how far one wants to project the heated excitement over the rapid progress of computer technology to an existential principle and evolutionary trend that ranks up there with sea creatures crawling onto the land.

Nevertheless, the strange thought that our own technology may one day awaken to consciousness and experience a subjective reality replete with emotions and survival instincts (not to mention intellectual and artistic ambitions) is one that has gained resonance over the past few decades. If the emergence of spiritual machines is even a remote possibility, it will make for fascinating reading as science fiction starts exploring more deeply the nature of machine art and culture, perhaps even machine psychology. Will we live at peace with our silicon children, or will we, like Dr. Frankenstein, fail to accept the God-like responsibilities that come with fathering a new species (Shelley, 2003)? Will we abandon our machines to slavery, fail to recognize their ethical rights as living beings and attempt to destroy them to prevent them from sharing our resources? Or is Kurzweil’s vision just a radical form of techno-fetishism, another primal human attempt to discover the keys to immortality using the tools of the present day?

As the Egyptians developed the technology of mumification to attempt to discover the immortality of the soul, so, too, are we building the computer chips and the android bodies that will carry about our consciousness? Perhaps the whole notion of computers awakening to consciousness is as devotedly mythological and religious as any belief lifted from the Egyptian Book of the Dead; and the belief that such a thing is even possible—firmly rooted in science and therefore more authentic than other immortality myths—can very well be the religious faith and spiritual prejudice of our modern scientific world fulfilling the same need to alleviate the anxieties about death and to promise immortality, where we are either floating like angels through some heavenly landscape or roaming the earth in an android body, half human and half computer, with a consciousness that can be removed like a floppy disk. It may very well be possible that future generations will look back at our junked computers, abandoned mainframes and attempts to create AI-driven android bodies with the same despair that we experience when we see the rotting and shriveled corpses of the Pharaohs who thought they were hitching a ride onto Ra’s Eternal Chariot.

The greatness of Philip K. Dick’s novel is not just that it addressed the dangers of our own AI machines, but that it goes even further to explore the fears and anxieties at the heart of our quest to overcome our mortality. It is not just the Nexus-6 that are threatening our humanity, but our own failure to feel as human beings for the suffering of others and to be able to gaze into the abyss of our own death. *Blade Runner* (Deeley), the 1982 movie adapted from *Do Androids Dream of Electric Sheep?* (Dick, 1968), added a touching finale to a violent and intense story: a moment in which the last remaining android spares Deckard his life after feeling empathy for him, a feeling provoked by the android’s recognition that human beings, too, fear death and struggle to overcome it. For one single moment, the dying android looks at Deckard, his destroyer, and smiles. In that smile is the plentitude of shared experience. A silicon machine and a carbon machine, both fearing their own death, both humbled at the threshold of that death.

So perhaps empathy is the key to our understanding of what it means to be human, and perhaps, in the end, if our machines do indeed wake up to their own consciousness, it will be our key to understanding them. Either way, we are faced with a very ponderous dilemma, one that has already generated vast anxiety as reflected in our science fiction: the fact that we have already begun to merge with our machines. We simply cannot shut off our computers without bringing human society to a violent end. It seems that by turning off our machines, we effectively turn off ourselves.

As Ray Kurzweil points out, even 30 years ago wasn’t the case.

Now it is.

References


Beethoven, L.V. (1824). Symphony No. 9 in D Minor, opus 125 [Synphonic composition]. Vienna, Austria.


Machine Learning in 1800s. As a concrete example, let's say we want to estimate pizza price using Linear Regression. We think that pizza radius, a number of salami slices and a number of tomato slices could affect the price (i.e. we've defined 3 features: radius, salami count, tomato count). We can write it this way: $y = \text{computer} + \text{graphics} + \text{photoshop} - \text{kidney} - \text{treatment} - \text{pill}$; if $y > 0$ then text is about computer graphics; if it is less than zero then we have a medical document. Systems based on Machine Learning are notably hard to debug; one of the reasons is that they often can adapt to such software bugs - usually it just costs us a small quality drop.

THE COMPUTER MOUSE When is a mouse not a mouse? When it sits on your desk and controls your on-screen cursor. The computer mouse is a pointing device with a flat bottom, a shaped top with buttonson and a cable connecting the mouse to the computer. The mouse moves on the surface of the desk, and the cursor copies this movement on the screen. The first computer mouse had wheels that made contact with the working surface.